EXAMINING THE RELATIONS OF MICRO- AND MACRO-LANGUAGE SKILLS TO
PERSUASIVE WRITING USING EXPLORATORY AND CONFIRMATORY FACTOR
ANALYSIS WITH APPLICATION OF GENERATIVE ARTIFICIAL
INTELLIGENCE-DERIVED FEEDBACK

By

Heqiao Wang

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Special Education—Doctor of Philosophy

2025

ABSTRACT

This dissertation explores the multifaceted development of persuasive writing skills among secondary school students, with an emphasis on fostering reasoning and argumentation skills essential for targeted writing tasks. Student-constructed essays serve as valuable instruments for assessing scientific literacy and higher-order thinking; however, their evaluation involves many complexities and is susceptible to potential biases and various sources of measurement error. This study addressed two primary objectives. First, it identified key features influencing persuasive essay quality by utilizing product-oriented measures to analyze both microstructural and macrostructural dimensions of writing. A latent structure of writing assessment was established, and measurement invariance was tested across students with differing special education statuses. Second, building on insights from this foundational investigation, automated feedback prompts were developed and implemented using GPT, facilitating AI-based scoring and feedback mechanisms. This research underscores the relevance of factor-analytic findings in addressing gaps related to predictive models and AI-driven content-generation tools, ultimately supporting personalized learning and adaptive written feedback. The implications of this study align with AI's expanding role in education, offering strategic insights into maximizing AI's utility for enhancing educational equity and instructional effectiveness. A promising approach involves leveraging factorial models to inform generative AI in delivering tailored feedback, thereby enhancing students' writing proficiency and overall performance.

To my beloved parents and grandparents, whose unwavering love and support have been my anchor throughout this journey. This accomplishment is as much yours as it is mine.

ACKNOWLEDGEMENTS

For years, I've rehearsed this moment in my dreams – the chance to express deepest gratitude to those who made this dissertation possible. Now that the time has come, I find myself at a loss for words, knowing that no acknowledgment could ever fully capture the support, wisdom, and kindness I've received along the way.

First and foremost, I extend my deepest gratitude to my advisor and committee chair, Dr. Gary Troia, whose guidance and support have shaped every stage of my doctoral journey. Your rigorous approach to science and writing not only fostered my academic growth but also ignited my passion for scholarly inquiry. During moments when I was still finding my voice as a researcher, your belief in my potential gave me the courage to push. What I learned from you during the past years provided exactly what I needed to develop as a scholar in education. This dissertation would not exist without your belief in me.

My heartfelt thanks also go to Dr. Kevin Haudek, my RA supervisor and committee member, for your invaluable support and guidance in helping me navigate the intersections of my research with broader fields. Working with you has been an intellectually rewarding journey – your insightful perspectives have strengthened my work in AI-enhanced educational assessment and profoundly shaped my research trajectory. I am deeply grateful for the opportunity to learn and grow under your guidance.

I am also deeply thankful to my other committee members, Dr. Kylie Gorney and Dr. Troy Mariage, for your keen interest in my work and for your insightful feedback, statistical expertise, and thoughtful contributions to this dissertation. A special thanks to Dr. Mariage – your courses and practical wisdom taught me how to bridge the gap between theory and practice, and I will always treasure collaborating with you on our undergraduate courses at MSU.

To my MSU community – I am profoundly grateful to the faculty I had the honor of TAing for, collaborating with on projects, and learning knowledge from (Dr. Adrea Truckenmiller, Dr. Eunsoo Cho, Dr. Spyros Konstantopoulos); to my incredible cohorts from special education and other programs (Dr. Cherish Sarmiento, Dr. Tingting Li, Eunha Kim) for their camaraderie; to my NGCI colleagues; to my statistical friends (Dr. Bixi Zhang, Shimeng Dai) for their expertise; and to all my friend, both near and far. Living in Lansing while pursuing my doctorate has been a joy, thanks to this wonderful community that surrounded me.

Above all, to my family – my eternal thanks to my father, Jingyang Wang, and my mother, Xinchun Wang, for being the wind beneath my wings. You are the unwavering support that has enabled me to soar. You generously invested in my education from my childhood through my master's degree across the ocean. You've given everything you had, even the things you never had the chance to experience yourself, to ensure I could become a better version of myself. To my grandparents, Wei Wang and Huixia Wang, thank you for surrounding my childhood with warmth and joy that its glow still guides me today. The happiness I experienced in my hometown during my early years helped me become the optimistic person I am now, giving me the strength to face and overcome the challenges I encountered along the way. To my boyfriend, Xiaohu Lu, thank you for standing by myside throughout this doctoral journey, both physically and emotionally. Your intelligence, perseverance, and dedication to research have been a constant source of inspiration. I am especially grateful for your patience and for offering comfort during moments of rejection, frustration, and uncertainty. Your support has meant the world to me, and I am confident that a brighter future awaits us.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

Effective writing is a foundational and important skill that individuals regularly employ across diverse educational and professional contexts (Attard, 2012; Coker Jr. et al., 2018; Fitzgerald & Shanahan, 2000; Graham & Alves, 2021; Kent & Wanzek, 2016; Troia, 2014). According to established writing models such as the Cognitive Process Theory of Writing (Flower & Hayes, 1981), the Simple View of Writing (Berninger et al., 2002), the Not-So-Simple View of Writing (Berninger & Winn, 2006), and the Direct and Indirect Effects Model of Writing (Y.-S. G. Kim & Graham, 2022), writing proficiency involves a coordinated integration of basic component skills (e.g., grammar, spelling, sentence structure), cognitive thinking processes (e.g., transcription, ideation, interpretation), executive functions (e.g., attention, goal setting, self-regulating), various knowledge domains (e.g., text structure, content, genre), and motivational attributes (e.g., self-efficacy, goal orientation, task interest and value) that writers engage during writing. A competent writer typically begins by acquiring essential skills such as solid transcription and ideation (often around grades 3 or 4), and then advances to more strategic writing by developing metacognitive abilities and gaining increased knowledge, often through instructional practices used at the secondary level. Empirically measuring writing skills presents significant challenges due to the complex nature of cognitive processes (e.g., information processing, problem solving) and psychological factors (e.g., affective stance, emotional regulation) involved. These challenges are particularly pronounced when analyzing open-ended responses and constructed essays, where biases may emerge from multiple sources of measurement error, including raters' judgments, the backgrounds of writers and raters, individual characteristics of writers, writing prompts employed, and the writing register or genre (Wang & Troia, 2023b). Additionally, districts and schools are increasingly expected to utilize assessment

data to monitor students' progress and interpret their responses to core and tiered instruction (Bondie et al., 2019). Researchers also call for providing personalized instructional practices based on the identification of students' strengths and weaknesses through assessment (Butterfuss et al., 2022; Philippakos & FitzPatrick, 2018; Troia et al., 2022).

Analyzing students' writing performance typically involves two major approaches to capture the multifaceted features of writing. The first approach involves examining students' writing-related processes and abilities, such as handwriting or typing fluency, spelling, decoding and word reading, working memory, background and genre knowledge, and motivation. These measures are usually collected through standardized and researcher-designed tasks, tests, and surveys, which help observe and understand students' performance on these features and their application in varied writing contexts. These foundational processes and abilities are critical as they may influence both the quality and quantity of writing across various tasks. The second approach focuses on analyzing students' prompted writing products by assigning different writing tasks, genres, purposes, or scenarios. Researchers then use holistic or analytic scoring schemes to quantitatively assess the resultant drafts. Raters and researchers typically examine either microstructural or macrostructural elements, or both, of an essay draft, including features such as linguistic and rhetorical components, content, structure, tone, and style. This approach provides insight into students' written performance, emphasizing a snapshot of the final product rather than the processes activated during the writing phase.

Today, researchers are exploring the automation of diagnostic processes within formative writing assessments using learning analytics tools and techniques. This involves progress monitoring on writing abilities (e.g., typing fluency: Truckenmiller et al., 2019; keystroke logging: Leijten & Van Waes, 2013; knowledge to marshal text evidence: Correnti et al., 2020),

quantifying text-based features (e.g., Crossley et al., 2014; McCaffrey et al., 2022), generating macrostructural features (e.g., Edwards, 2003), identifying topical patterns (e.g., Kuzi et al., 2019), and observing emotional and behavioral engagement in writing activities (e.g., Liu et al., 2018; Roscoe et al., 2017), with the aim of enhancing evaluative efficiency in writing assessments. The predictive aspect of writing scores is not the endpoint of the automated process; rather, the understanding gained from the prediction holds empirical significance (Wang & Troia, 2023b). However, there is a lack of research studies on how the prediction process can be further developed for other educational purposes, such as personalizing students' learning and providing feedback based on their current writing performance. Generative AI has the potential to facilitate this process.

1.1 PRESENT STUDY

Many students struggle with persuasive writing. It is essential to conduct a meaningful and informative assessment of persuasive writing performance to understand students' argumentative capacity. While research has explored various factors influencing the quantity and quality of persuasive writing – such as linguistic/rhetorical measures (Jo, 2022), argumentative structure and substance (P. Stapleton & Wu, 2015), and cohesion and coherence (Andreev & Uccelli, 2024) – there remains a significant gap in understanding the internal structural relationships among these factors, which is crucial for a comprehensive evaluation of argumentative capacity. This study aims to address this gap by introducing a proof of concept for assessing persuasive writing through product-oriented measures, focusing on microstructural and macrostructural levels of performance. Moreover, AI's ongoing prominence in our zeitgeist emphasizes the importance of maximizing its utility in education. Derived factorial models have potential to guide Generative AI in providing written feedback and enhancing students'

performance, as discussed by Steiss et al. (2024) and Meyer et al. (2024). Specifically, this study seeks to answer three research questions:

1) What textual attributes serve as optimal indicators of persuasive essay quality in secondary school students?

2) To what extent do secondary students with different special needs status (i.e., students with versus without an Individualized Education Plan [IEP]) exhibit significant differences in their holistic writing scores across latent writing attributes?

3) Do essays revised by GPT, a Generative AI application, utilizing prompts derived from factor analysis, demonstrate enhanced performance compared to the original essays written by students?

1.2 ORGANIZATION OF THE DISSERTATION

Following this introductory chapter, Chapter Two presents a comprehensive overview of the literature pertaining to writing performance, with a specific focus on persuasive writing at the secondary education level. The chapter establishes the theoretical frameworks or perspectives that underpin the study's measurement and analysis. Through synthesizing empirical literature, the chapter identifies specific linguistic and structural features that characterize persuasive writing. Furthermore, it justifies the study's significance by addressing gaps or problems identified in the extant literature. Lastly, the chapter provides a rationale for the study by highlighting its educational significance and potential contributions to the field of writing education.

Chapter Three provides a detailed examination of the study's methodology. This study primarily employs a secondary data analysis of an extant dataset (the PERSUADE 2.0 corpus) approach grounded in quantitative research methods. The chapter begins by outlining the study

purposes and restating the three research questions noted above. It then describes the dataset and its variables, including derived variables at both microstructural and macrostructural levels, which are employed to analyze persuasive writing samples from secondary students in the corpus. Furthermore, the chapter elaborates on the research design employed to address the three research questions.

Chapter Four organizes and reports the study's major findings. This includes descriptive results from the exploratory data analysis and the complete results addressing each of the research questions proposed. This chapter also synthesizes and discusses the results in the context of the study's research questions and theoretical background for each research question.

Chapter Five presents practical implications, concluding statements, and outlines future directions. The implications are intended to guide scholars and educators in applying the study's findings to real-world settings. They offer strategies for integrating human intelligence with automated methods to design more effective writing prompts and essay evaluation tools that facilitate feedback and support accurate quality assessment. The study's limitations are also noted, and recommendations for future research based on the overall results of the dissertation are included.

Chapter 2: Review Of Literature

As stated in the preceding chapter, the primary objective of this study is to undertake exploratory and confirmatory factor analyses on **M**icrostructural and **M**acrostructural **F**eatures that underpin **P**ersuasive written composition (MMFP) and their relations to overall quality of persuasive writing. To achieve this aim, the study utilized a large-scale corpus comprising persuasive essays written by middle and high school students with diverse sociodemographic traits. In addition, as a secondary objective, the study sought to validate the practical application of insights derived from MMFP by using them to provide constructive written feedback to students using Generative AI (GenAI) systems, thereby potentially enhancing their persuasive writing skills and informing future instructional decisions related to revision and editing.

The initial sections of this chapter offer background information that emphasizes the importance of writing in academic and professional contexts and illustrates students' current writing performance broadly across all grade levels, but specifically at the secondary level, with a focus on persuasive writing. Subsequently, the chapter also discusses various evaluation concerns related to persuasive writing, informed by prior research that includes supporting theories, empirical findings, and measurement considerations. Lastly, the chapter concludes by discussing research gaps, the study's approach to addressing these identified gaps, and the educational significance of the methodology employed in this investigation.

2.1 THE ROLE OF WRITING

2.1.1 General Significance of Writing

In contemporary society, the written word functions as a versatile tool for attaining varied social and educational objectives (Graham, 2006; Klimova, 2013). For example, correspondence mediums such as letters, postcards, and emails can facilitate communication and foster

interpersonal bonds when physical distance separates individuals. The practice of maintaining a personal diary and engaging in self-reflection is recognized for its capacity to promote contemplation on self-identity, develop short/long-term goals, and alleviate feelings of loneliness, thereby yielding psychological and physiological benefits (Smyth, 1998). In professional spheres, composing well-structured reports and summaries may contribute to effective information dissemination and documentation of project outcomes within teams and to organizational leadership. Additionally, writing is acknowledged for its capability to influence others' perceptions, emotions, and beliefs (Graham et al., 2012) and serves broader social goals such as shaping public opinion (Aldisert, 2009), fostering empathy (Dhurandhar, 2009), and advocating for social justice (Singh, 2011).

Within the scope of this study, focusing on the educational domain necessitates an exploration of the pivotal role that writing assumes in academic success. In the context of K-12 English Language Arts (ELA) education, students undergo a developmental trajectory marked by the acquisition of increasingly proficient and sophisticated writing skills, coupled with cognitive and metacognitive strategies, as they progress through grades (Graham & Harris, 2010; G. A. Troia et al., 2013; Wang & Troia, 2023a). At the early stages of learning to write, narrative writing holds a significant place in the early elementary school curriculum, largely due to its resonance with oral language expression (Spencer & Petersen, 2018). Students engage in crafting imaginative worlds, constructing plots, and developing characters. These practices have been empirically demonstrated to enhance students' capacities for creative thinking (Eser & Ayaz, 2021) and storytelling (Rambe, 2017). Narrative writing also serves as a nexus for reinforcing basic component skills, including transcription, vocabulary and grammar usage, and sentence construction (Olinghouse, 2008; Salas & Silvente, 2020; Wiliana & Djajanegara, 2019).

7

Proficiency in these foundational areas holds considerable implications for overall writing quality and establish a robust foundation for the subsequent acquisition of skills in more complex writing genres that require advanced writing skills (Deane et al., 2008; Puranik & Lonigan, 2014).

From grades 3 and 4 onwards, students are expected to read and write informational text across diverse content areas, and this expectation intensifies as they advance through school (Jeong et al., 2010). The significance of the informational genre is underscored by its prevalence in many state and national writing assessments. Research indicates that the proportion of informational text featured in standardized tests can be as substantial as 70% to 80% (see Palumbo & Sanacore, 2009). Students with limited exposure to informational text are prone to achieving lower scores on these standardized assessments (Heider, 2009). Prior studies reveal a positive correlation between students' proficiency in writing informational text (including those who have received targeted informational writing interventions) and enhanced performance in multiple learning dimensions (Graham & Perin, 2007; Graham et al., 2012). This positive correlation may be attributed to the fact that informational writing can lead to a better understanding of content area concepts (Parson, 2013), increased knowledge about key topics (Taboada & Guthrie, 2006), and improved information processing and abstraction (Fox, 2009).

Persuasive writing is another essential writing genre systematically incorporated into K-12 education. The National Assessment of Educational Progress (NAEP) writing report card (National Center for Education Statistics [NCES], 2012) underscores that students often encounter heightened challenges with persuasive writing more so than narrative, descriptive, or expository registers. Within this genre of writing, students are tasked with explaining complex and interdisciplinary concepts by incorporating specific, relevant details, substantiating their

claims, and effectively convincing their intended audience to align with their stance on a topic. By the end of twelfth grade, students are anticipated not only to adhere to the foundational aspects of content standards in the Common Core State Standards (CCSS; National Governors Association Center for Best Practices [NGACBP] & Council of Chief State School Officers [CCSSO], 2010) but also to extend their persuasive composition capabilities, including the ability to differentiate claims for alternate or opposing viewpoints, and to clarify the relationships between claims and reasons, reasons and evidence, and claims and counterclaims. This increased complexity renders persuasive essays more demanding, as outlined in the CCSS in ELA for persuasive writing. Mastery in persuasive writing guides students towards nuanced argumentation (Brockman, 2020), evidence-based reasoning (Hemberger et al., 2017), and effective communication of ideas (F. I. A. Aziz & Ahmad, 2017), thereby fostering a comprehensive and higher-order skills set in written expression.

The act of writing is often considered synonymous with the act of thinking. It is a cognitive process, which necessitates the application of analytical thinking skills to creatively and critically organize ideas (Flower & Hayes, 1981; Grimberg & Hand, 2009; Menary, 2007). Previous research studies have illuminated various educational and psychological benefits associated with the act of writing. In the educational realm, as noted earlier, practicing writing across different genres has been identified as a means to enhance students' learning in multifaceted ways. The efficacy of writing to learn is increased when students are guided to employ specific cognitive and metacognitive strategies of self-regulated learning (Fry & Villagomez, 2012; Hübner et al., 2010). Furthermore, aligning writing tasks with students' preferred writing approaches can enhance the effectiveness of this process (Kieft et al., 2008). The practice of writing to learn supports students' active learning, retention, and writing

9

development (Fry & Villagomez, 2012). For instance, research indicates that college students, when engaged in process-oriented writing activities, exhibit a higher level of comprehension of scientific concepts and greater knowledge acquisition compared to controls using multiple-choice assessments to evaluate their understanding of the source materials (Royse et al., 2024).

Moreover, the act of writing proves particularly beneficial in bolstering students' learning outcomes within science, technology, engineering, and math (STEM) disciplines, as it serves as a vehicle for thought, reasoning, and knowledge-in-use (Boscolo & Mason, 2001; McNeill & Krajcik, 2009). Klein (2006) observed that non-STEM majors at the postsecondary level showed greater posttest transfer of scientific concepts when they processed and conveyed new information through writing rather than through verbal expression. However, research comparing students' oral and written expression of scientific concepts remains limited, and the reasons for the differences between these modalities remain underexplored. Writing often involves the use of more precise and sophisticated vocabulary than oral expression, which may explain why writing-to-learn activities are frequently considered more effective for assessing students' conceptual understanding in disciplinary education (Chen et al., 2023; Royse et al., 2024; Visser et al., 2018). This potential makes the textual analysis of vocabulary in written form particularly pertinent to this study, as vocabulary plays a critical role in carrying meaning at the secondary school level. This highlights the importance of written expression as a valuable tool for assessing student knowledge.

From a psychological perspective, writing can promote self-concept and self-efficacy. For example, in descriptive writing, where students structure their experiences and derive meaning from significant event, the writing process enables their integration of their personal experiences into their self-schema, contributing to the development of more positive self-perceptions

(Graybeal et al., 2002). In turn, students with enhanced academic self-concept following writing tend to achieve greater academic success in subsequent academic years (Muijs, 2011; Pajares et al., 1999). This correlation is attributed in part to a clearer self-awareness regarding writing strengths and weaknesses, which empowers writers to identify areas for improvement. Writing can enhance self-regulation and bolster individuals' control over challenging thoughts and emotions. It allows individuals to actively observe, monitor, and assess their emotional expression and regulation (Schmitz & Perels, 2011). The resulting sense of control over emotions directly contributes to improved well-being and a reduction in negative emotions (C. M. Stapleton et al., 2021).

2.1.2 Academic Significance of Secondary Writing

Despite the significant role of writing in educational, psychological, and social domains, a substantial number of students in the United States graduate from secondary schools without attaining proficiency in writing (Graham et al., 2014). Only 27 % of grade 12 students demonstrated performance at or above the "proficient" level in writing according to NAEP (NCES, 2012). This indicates a widespread deficiency across the nation in constructing written responses that effectively achieve the communicative objectives of writing, with "proficient" writing characterized by well-organized and coherent text coupled with appropriate transitions and diverse sentence constructions (S. A. Crossley & McNamara, 2016). In addition, half of twelfth grade learners grapple with rudimentary aspects of writing, such as employing detailed and factual descriptions, making appropriate lexical choices, and utilizing varied sentence structures (e.g., Wang & Troia, 2023a). To understand these concerns, it is imperative to elucidate the educational significance of writing at the secondary level.

2.1.2.1 Academic Writing as a Discipline-Specific Skill

Guided by the belief that the responsibility for instructing secondary academic writing should be a collaborative effort involving both ELA teachers and educators from other academic fields (Russell, 1997), the Writing Across the Curriculum Clearinghouse serves as a publishing collaborative that has spurred a reevaluation of secondary student academic writing within educational communities. This reconceptualization includes acknowledging students' literacy and language experiences beyond the classroom, incorporating the principles of Writing in the Disciplines (WID; Blumner & Childers, 2016). WID is dedicated to developing socially mediated communication skills and genre knowledge specific to individual academic disciplines (Broadhead, 1999). In fact, there is a noticeable absence of writing or writing instruction in the typical ELA middle school classroom (Applebee & Langer, 2015; Graham & Perin, 2007), but there has been an intensified focus on writing instruction in other content areas such as science-related subjects (Miller et al., 2016). Aligned with CCSS objectives, writing can be considered as an instrument to facilitate learning across varied content areas, which signifies a significant shift in policy and practice within secondary schools in the U.S. The diversity in disciplinary practices mandates distinct modes of written communication at the secondary level, such as technical writing in the sciences, persuasive writing in business, and argumentative writing in literary analysis (Ezza et al., 2020). To conclude, the secondary school environment is characteristically "discipline-driven" and "discipline-delineated" (Miller et al., 2016). High school students are now required to use writing as a tool for analyzing and reflecting on information in language arts, social studies, science, and various technical subjects.

A substantial body of literature has revealed the instructional practices of writing across various disciplinary areas in middle and high school (e.g., Anders & Guzzetti, 2020; Graham et

al., 2020; Shemwell, 2020). Drew and colleagues' (2017) study provides valuable insights into the instructional strategies employed in disciplinary classes, specifically focusing on grades 6 to 12 science classes. Through qualitative analysis of teacher surveys, the study reveals that science teachers recognize the alignment of writing with the broader objectives of science education. They intentionally choose to integrate writing into their science classes, teaching students how to create scientific texts and to acquire and utilize scientific vocabulary, which ultimately supports knowledge building and application. The study further emphasizes that teachers who assign writing tasks that allow students to analyze and synthesize information while employing discipline-specific genres are more likely to promote deep learning in their students. This approach not only enhances students' understanding but also empowers them to contribute to effective scientific communication, which is also consistent with the overarching goal of science education as outlined by National Research Council (NRC, 2012). Lastly, the study identifies evidence-based practices for teaching writing during secondary science classes by demonstrating the efficiency of incorporating writing strategy instruction within inquiry-based pedagogy to support learners of all proficiency levels.

Graham and colleagues (2013) offered a comprehensive national overview of middle-school writing instruction, drawing from participating teachers' self-reports regarding their preparedness to teach writing, beliefs about teaching responsibilities, utilization of evidence-based writing instruction practices, assessment methods, incorporation of technology, and adaptations for struggling writers. A significant finding from this national survey indicates that middle school teachers perceive the teaching of writing as both a personal and shared responsibility, despite many educators across disciplines lacking sufficient preparation to teach writing and facing substantial constraints on instructional time dedicated to writing. Within the

limited instructional time available, teachers focus on employing writing activities to enhance students' learning and instruct them on techniques for summarizing material they read. On a monthly basis, teachers encourage students to emulate models of proficient writing, provide instruction on fundamental writing skills, evaluate students' writing using rubrics and other assessment tools, and teach strategies for planning, revising, and crafting paragraphs. This study offers insights into the deficiencies of writing instruction in middle school, highlighting issues such as inadequate time allocated for teaching writing and limited opportunities for students to engage in writing activities. The study suggests that addressing these gaps requires collaborative efforts with educators from various disciplines.

Academic writing at the secondary level serves as a valuable foundation for content learning and is intrinsically linked to subsequent postsecondary pursuits. The overarching objective of disciplinary literacy is to equip students with the necessary skills for sophisticated literacy demands in college and careers (G. A. Troia & Maddox, 2004). This preparation is accomplished through focused instruction on discipline-specific literacy strategies within core content-rich areas, such as mathematics, history/social studies, and science, at both the secondary and postsecondary levels (Fang & Schleppegrell, 2010; Shanahan & Shanahan, 2008).Teachers in various disciplines who adopt WID methodologies, whether intentionally or implicitly, impart to their secondary students the expectations and nuances of writing in a collegiate context, which can better facilitate students' smooth transition to postsecondary writing expectations and tasks. The strategies, knowledge, and skills derived from secondary academic writing and writing instruction can be effectively transferred to ensure success in postsecondary writing endeavors (WWC, 2016).

2.1.2.2 Writing Using Cognitive and Metacognitive Abilities

In alignment with the CCSS for ELA standards, secondary school writing necessitates students to "write like specialists" (Dressen-Hammouda, 2008). In addition to being discipline-driven, secondary school writing also requires students to master more sophisticated and higher-order writing skills, strategies, and conventions in comparison to elementary school-aged writing, which often concentrates on basic elements such as spelling, handwriting, grammar, and essential text elements and basic genres. The complexity of writing increases for secondary school-aged students, demanding not only cognitive resources for organizing, storing, and activating knowledge and skills in the composition process (Shen & Troia, 2018; Weinstein & Hume, 1998) but also necessitating adequate metacognitive abilities and strategies (Graham & Harris, 2010; Yamson & Borong, 2022). These metacognitive skills enable students to monitor their writing-related thoughts, emotions, and behaviors, maintain a positive attitude toward writing, and optimally utilize cognitive processes to achieve learning objectives (G. Troia, 2014).

In secondary classrooms, writing tasks often involve limited analysis, interpretation, and composition (Applebee & Langer, 2015). For example, students frequently encounter writing assignments such as worksheets or brief composing tasks that require reflection on source texts, followed by creating mental representations of the texts and producing responses or summaries based on these cognitive processes (Cer, 2019). More advanced-level writing in secondary education demands that students engage in complex cognitive activities such as estimating the needs of one's audience, setting long- and short-term rhetorical and personal goals, self-monitoring writing processes and performance over time, and self-evaluating outcomes in comparison with established goals (Graham & Harris, 2010). These cognitive and metacognitive activities not only assist secondary education students in crafting high-quality texts but also help

writers supervise and correct written errors, enhance the overall learning process, and contribute to the development and regulation of learners' awareness at rhetorical and cognitive levels for writing (Ramadhanti & Yanda, 2021).

In the meta-analysis conducted by Dignath and Büttner (2008), it was found that metacognitive knowledge and strategies, such as self-regulated learning, can be effectively nurtured at both primary and secondary school levels. The overall effect size (ES) of interventions/programs on students' writing performance at the secondary school level (ES = 0.71) was observed to be slightly higher than that at the primary school level (ES = 0.68). This may be attributed to the fact that children entering primary school typically exhibit limited reflection and control over their learning compared to their counterparts entering secondary school (Paris & Newman, 1990). Research on metacognitive development indicates that younger or inexperienced student writers often encounter challenges in utilizing metacognitive strategies. This is because they may not have sufficient cognitive capacity available to employ additional strategies alongside the demanding tasks of reading or writing (Alexander et al., 1998). In contrast, older or mature students who have automated the processes of reading and writing have more cognitive capacity available for metacognitive activity. Consequently, they can derive greater benefits from strategy training in this context.

Explicit engagement in cognitive and metacognitive activity through writing instruction at the secondary school level yields valuable benefits. It enables students to contemplate their individual learning characteristics and aids them in mastering content (Conley, 2014). Through the entire process, students can assess their knowledge and skills in specific areas, develop a repertoire of strategies to acquire knowledge, and discern appropriate actions for various academic tasks (Bürgler et al., 2021; Hartman, 2001). These components, encompassing self-

knowledge, reflective thinking, planning and organizing, employing effective strategies, and evaluating written products, which all involve metacognition, ultimately contributing to students' success in their postsecondary pursuits (Bauer, 2014; Mytkowicz et al., 2014).

2.2 AN OVERVIEW OF STUDENT WRITING

2.2.1 Current State of Writing Performance

With the widespread adoption of the CCSS across the nation, most states have developed or embraced new writing assessments for elementary through high school students (Kelly-Riley, 2017). The rapid expansion and evolution of educational testing in the United States over the past several decades have been significantly influenced by university initiatives aimed at shaping secondary-level curriculum to better prepare students for postsecondary-level work (Ramirez et al., 2018). Students demonstrate significant variability in their written composition performance, with a substantial majority (on average, 73%) performing below proficiency standards across elementary, middle, and high school levels (Truckenmiller et al., 2021). Building on the earlier discussion regarding the role of writing as an indispensable component of the K-12 curriculum, it becomes crucial to assess the current state of students' writing performance through state and national standardized assessments.

2.2.1.1 State Testing

Supported by the federal Race to the Top initiative (U.S. Department of Education, 2009), two state-led consortia, namely the Smarter Balanced Assessment Consortium (SBAC) and the Partnership for Assessment of Readiness for College and Careers (PARCC), launched their novel assessments in the spring of 2015, impacting roughly half of the U.S. Even in states not aligned with these consortia, new writing assessments were formulated to adhere to the CCSS and vie for funding from the Race to the Top program. Over forty-one states include a writing component in

their ELA assessments (Jeffery, 2009). In twenty of these states, students must pass the state test to graduate from high school (Kober et al., 2011). Nevertheless, numerous high school students encounter challenges in writing, particularly those with disabilities.

In Michigan, the current statewide assessments primarily include the M-STEP (Michigan Student Test of Educational Progress), a summative assessment gauging students' knowledge and capabilities based on Michigan's CCSS-aligned academic standards. Additionally, the PSAT (Preliminary Scholastic Assessment Test) assesses eighth grade reading, writing, and math. MI-Access is another test aligned with Michigan's alternate content expectations, designed for students with significant cognitive disabilities, for whom the M-STEP, even with accommodations, is deemed inappropriate. Michigan's state assessment program aims to furnish districts and schools with information about students' proficiency based on the academic standards, aiding in the formulation of continuous improvement goals. These state assessments also offer teachers, parents, and other stakeholders' insights into individual students' knowledge and performance in key content areas, ensuring compliance with the *Every Student Succeeds Act* and the *Individuals with Disabilities Education Act*.

Table 2-1 presents the percentage of students across grades 3 to 8 falling into the various performance level categories of advanced, proficient, partially proficient, and not proficient for M-STEP and PSAT during the 2022-2023 school year. The M-STEP results reveal that only 20.3% to 27.1% of students in Michigan across grades 3 to 7 achieved the proficient level in ELA. Notably, 63.1% of seventh graders are partially or not proficient according to their ELA scores. For Grades 3 to 5, over 40% of students achieved advanced or proficient levels, but this figure decreases starting from grade 6, where less than 40% of students achieved the advanced or

proficient levels. The PSAT results also indicate that 40.3% of eighth-grade students did not achieve advanced or proficient levels according to their ELA scores.

Table 2-1 Grades 3-8 ELA Michigan Testing Performance Level in 2022-2023

Assessment Name	Grade Level Content	Students Advanced/ Proficient	Advanced	Proficient	Partially Proficient	Not Proficient	Partially/Not Proficient	Number Assessed
M-STEP	3rd Grade	40.9%	20.6%	20.3%	24.5%	34.6%	59.1%	98,715
M-STEP	4th Grade	44.3%	22.7%	21.6%	20.0%	35.6%	55.7%	97,894
M-STEP	5th Grade	43.9%	17.4%	26.5%	21.3%	34.8%	56.1%	98,403
M-STEP	6th Grade	37.5%	11.6%	25.9%	26.7%	35.8%	62.5%	99,114
M-STEP	7th Grade	36.9%	9.8%	27.1%	28.1%	35.0%	63.1%	98,211
PSAT	8th Grade	59.7%	37.7%	22.0%	15.1%	25.1%	40.3%	98,932

2.2.1.2 National Testing

The NAEP stands as the largest nationally representative assessment of various academic subject areas in the United States (Mo & Troia, 2017a). Enhancing the generalizability of NAEP results and providing actionable policy implications would significantly augment the value of NAEP findings (Williamson, 2006). Specifically, NAEP gauges the writing proficiency of U.S. students by administering assessments to sample groups representative of the nation's student population. The most recent available data on students' writing performance is from the 2011 school year, sourced from the U.S. Department of Education, Institute of Education Sciences, National Center for Education for Writing Assessment. The 2011 writing assessment marked the inception of the first NAEP computer-based writing assessment, developed under a new NAEP writing framework (National Assessment Governing Board, 2011) where the substantial role of computers in the writing process was acknowledged. Students were randomly assigned two writing tasks, each requiring 30 minutes for task completion. For eighth-grade students, 35% of prompts focused on writing persuasive essays, 35% on explanatory essays, and 30% on narrative essays. Grade 12 students were assigned 40% of prompts for persuasive essays, 40% for explanatory essays, and 20% for narrative essays. Furthermore, students were expected to engage

in various tasks for diverse audiences and participate in processes of generating, revising, and editing.

The results of the 2011 NAEP writing assessment for eighth and twelfth graders are detailed in Table 2-2. For eighth graders, 20% scored below basic, 54% were at or above basic and below proficient, 24% were at or above proficient and below advanced, and 3% achieved advanced. Among twelfth graders, 21% scored below basic, 52% were at or above basic and below proficient, 24% were at or above proficient and below advanced, and 3% achieved advanced. These findings indicate a significant lack of proficiency in writing among U.S. secondary school students, reflecting a decline when compared to the 2007 NAEP assessment results (Salahu-Din et al., 2008). This suggests that most secondary school-aged students are not adequately equipped with the necessary writing skills and knowledge for postsecondary education before high school graduation (Williamson, 2006).

Table 2-2 Average Scale Scores (Percentages) at Each Achievement Level for NAEP Writing Report in 2011

Grade Level	Average Scale Score	Below Basic	At Basic	At Proficient	At Advanced
8th Grade	150	20%	54%	24%	3%
12th Grade	150	21%	52%	24%	3%

2.2.2 Secondary Writing Difficulties

Through state and national assessment outcomes presented in the previous section, it is evident that a substantial portion of U.S. students do not attain proficiency in writing before graduating from high school. Many of these students, grappling to become more adept users of the discourses required for college-level classes, harbor a belief that they lack skills or talents needed to write well (National Commission on Writing, 2004). This negative perception often leads to a fear of failure in academic writing tasks, resulting in resistance to assignments– manifested through late submissions, incomplete work, resorting to plagiarism, or even failure or withdrawal from the class (Fernsten & Reda, 2011). It is crucial to comprehend the prevalent

20

writing challenges faced by students, particularly those at the secondary school level and, importantly, to understand the root causes of these writing difficulties.

Firstly, student writers, particularly those transitioning to junior high school, may still lack maturity in in basic writing skills (Graham et al., 2023). The rapid changes in writing demands at the secondary level may reveal gaps in these fundamental skills, impeding students' ability to effectively express their ideas. These gaps may include weak essential skills including transcription such as handwriting, typing, spelling, and punctuation, along with challenges in sentence construction arising from limited grammar knowledge, vocabulary, and exposure to diverse sentence structures (G. A. Troia et al., 2011). These factors significantly impact the readability of written text and overall writing performance on composition tasks (Graham et al., 1991). The transition from elementary to secondary school marks a pivotal phase where students are expected to refine their writing capabilities, and the adaptation to these elevated expectations can lead to initial struggles.

Secondly, aside from the basic component skills, students entering secondary school may frequently encounter challenges related to higher-order cognitive abilities that are essential for tackling the progressively complicated writing tasks demanded at this educational stage (Demetriou et al., 2020). According to Hayes' model of writing (1996), the individual writer's cognitive processes are intricately connected with elements of the task environment, encompassing both social factors (e.g., the audience, collaborators, cultural norms) and physical elements (the composing medium, text already produced, source materials). However, novice writers frequently overlook the importance of engaging in these cognitive writing processes (Salas & Silvente, 2020). Instead, they may rely on a retrieve-and-write process, wherein they compose solely by "generating or drawing from memory a relevant idea, write it down, and use

each preceding phrase or sentence to stimulate the next idea" (cited by Graham & Harris, 1997, p. 235). This approach simplifies writing tasks by eliminating the sophisticated advancement of rhetorical goals and simultaneously, minimizes the use of planning, monitoring, editing, revising, and other strategic behaviors (Graham & Harris, 1997). The neglect of these components may limit students' opportunities to evaluate their ideas and writing goals. Additionally, they are less likely to detect written errors and assess the overall organization of their text. In contrast, writers who actively engage in these writing processes typically employ more diverse and technical vocabulary in their compositions (Koutsoftas & Petersen, 2017) and demonstrate greater accuracy in their grammar usage (Mackie & Dockrell, 2004), among other positive attributes. Thirdly, students' knowledge base regarding writing and its genres, devices, and conventions can greatly impact their writing and present additional writing challenges (Graham & Harris, 2010). As students progress through middle and high school, writing tasks become more demanding with increased requirements related to writing conventions, prompts integrating reading sources, objectives for intended audiences, etc. (Deane et al., 2008). The process of crafting a coherent text involves the utilization of different types of knowledge (Hayes, 1996; Olinghouse & Graham, 2009; Saddler & Graham, 2007; Troia et al., 2022), including topic knowledge largely derived from source materials, prior knowledge that enables writers to extract and evaluate the information mentioned in the source materials, genre knowledge associated with writing intentions across genres/registers (e.g., narrative, explanatory, declarative, procedural writing knowledge), linguistic knowledge related to mechanics and discourse, and so forth. These knowledge sources collectively impact students' language use in terms of conventions, grammar, handwriting, and spelling, as well as guiding their adaptation of writing style, structure, and tone to meet the specific requirements of different writing tasks. There is a growing concern among

educational scholars that students at the secondary level may struggle to activate and apply their knowledge effectively in their writing endeavors (De La Paz & Graham, 2002; Graham et al., 2014; Trapman et al., 2018).

Fourthly, a lack of motivation to write represents another important factor that may undermine students' writing abilities. An array of writing motivation (including self-efficacy, goal orientation, task interest and value, attributions for outcomes) among secondary school students predict their writing competence (Klassen, 2002; Pajares & Johnson, 1994; Troia et al., 2012). However, students in middle school tend to have diminished self-efficacy beliefs and possess lower self-concepts than students in elementary school (Pajares & Valiante, 1999; 2001). This decline is linked to their perception of high school classes as emphasizing a performance goal orientation and having reduced confidence in their writing abilities, particularly concerning grammar, usage, and other mechanical skills appropriate to the complexity of assigned tasks (Pajares & Cheong, 2003). The decreasing self-efficacy and persistence toward writing at the secondary level renders students pessimistic about their capabilities to generate and organize ideas for writing, impeding their ability to transcribe ideas into sentences (Camfield, 2016; Pajares, 2003). Additionally, secondary students may lack the stamina and working memory capacity needed to both sustain their writing efforts and effectively correct errors in their papers.

Factors contributing to writing difficulties among secondary school-aged students can be attributed to both internal and external factors. Internally, psychological elements, such as stereotypes, can affect students from diverse cultural, racial, and educational backgrounds. These students may be conscious of stereotypes about themselves, such as being poor communicators, which can influence their writing behaviors and self-perceptions. For instance, students of color may internalize beliefs that they write less proficiently than their peers from dominant groups,

23

leading to poorer self-perceptions of their writing abilities (Abrahams, 1972). Children identified as at-risk in writing or with disabilities may experience diminished self-perceptions, exacerbated by comparisons with their peers within the classroom setting (Hamilton, 2011; Wright et al., 2021). This may deter their engagement with school writing tasks. Additionally, gender perceptions can impact students' writing experiences. Girls may perceive themselves as better writers than boys in their class or school, attributing this belief to stronger domain-specific self-concepts and self-efficacy for self-regulation (Pajares et al., 1999). Female writers also may exhibit lower apprehension toward writing tasks (Pajares et al., 1999). These internal factors contribute to the complex landscape of writing difficulties among secondary school students.

External factors also wield considerable influence on what challenges students encounter in writing. These factors involve aspects such as teaching styles employed by educators (e.g., Ramli et al., 2020), the quality of instructional materials (Wang & Troia, 2023a), the occurrence of corrected feedback (M. M. Nelson & Schunn, 2009), and the presence of peer support and scaffolding (Taheri & Nazmi, 2021). The absence of writing strategy instruction (Rodríguez-Málaga et al., 2021; Shen & Troia, 2018), writing process instruction (Graham & Sandmel, 2011; Troia & Olinghouse, 2013), and instructional scaffolds for motivation to write (Troia, 2002; Troia et al., 2012) can also hinder students' academic development. Additionally, the lack of guidance on utilizing digital writing tools can impede students' performance across various writing tasks (Ekholm et al., 2018). Incorporating these instructional components not only benefits students on diverse writing tasks (Troia et al., 2022) but also enables them to self-monitor and regulate their writing behaviors and outcomes (Graham & Harris, 2010). This, in turn, nurtures an awareness of their writing strengths and weaknesses, fostering metacognition as students engage in reflective thinking about their writing processes and products.

2.3 PERSUASIVE WRITING

2.3.1 Definition

The definitions of persuasive writing within rhetorical and educational discourse exhibit a degree of uniformity. For instance, Eemeren et al. (2001) characterize persuasion as "a verbal and social activity of reason, with the objective of enhancing (or diminishing) the acceptability of a contentious standpoint for the audience by presenting a set of propositions designed to justify (or refute) the standpoint before a rational judge" (see p. 5). Güneş (2016) posits that persuasive writing involves creating texts by specifying, elaborating, predicting, and justifying reasons in a manner acceptable to others. Purdue Online Writing Lab defines the persuasive essay as a written genre that requires students to explore a topic, gather, generate, and evaluate evidence, and succinctly establish a position on the subject. In the most prevalent U.S. English language arts standards, persuasive writing involves the writer's endeavor to convince or persuade the audience to adopt a specific point of view or take a particular action through the construction of logical arguments and a cohesive summary (CCSS for English Language Arts and Literacy in History/Social Studies Science and Technical Subjects, 2010).

The provided definitions highlight three key aspects of persuasive writing. First, composing a persuasive essay is a social activity by nature, involving dialogue among individuals (e.g., between audience and authors) who may hold divergent perspectives on an issue, aiming to convince each other by providing compelling evidence. Second, the presentation of a constellation of propositions suggests that arguments have a discernible structure and organization, as evidenced in much of the educational literature (Crammond, 1998; De La Paz et al., 2012; Durst et al., 1990; Uccelli et al., 2013). Third, constructing arguments within persuasive essays is an act of reasoning, and individuals of reason employ critical standards to

assess the acceptability of a standpoint. These critical standards may include criteria such as the incorporation of argumentative discourse elements, the writer's attentiveness to audience considerations, and the appropriateness of their argumentative strategies in relation to their objectives (Tindale, 2002). It is widely accepted that effectively defending arguments involves addressing critical questions about the pertinence of chosen argumentative strategies (Ferretti et al., 2009; Johnson, 2014; Walton, 2013).

2.3.2 Importance

Persuasive writing is a crucial skill for students to develop, as it is highly regarded in higher education contexts (F. I. A. Aziz & Ahmad, 2017). Additionally, persuasive writing is a challenging task that necessitates the use of sophisticated language to analyze, discuss, and address controversies in a manner that is clear, convincing, and respectful of different perspectives. In the educational context, as per the Michigan K-12 Standards for English Language Arts (Michigan Department of Education, 2010), which are based closely on the CCSS, the academic standards articulate the expected learning progressions for persuasive writing in Michigan. These standards serve as a deliberate framework guiding local curriculum development and emphasize the increasing proficiency expected of students in writing persuasively using various language dimensions as they advance through higher grade levels. These dimensions include adept mastery of vocabulary and syntax, development and organization of ideas, and engagement with progressively demanding content and sources.

For instance, when composing persuasive essays, grade 6 students in Michigan are required to build arguments supporting claims with clear reasons and relevant evidence by: (a) introducing claim(s) and organizing reasons and evidence clearly, (b) supporting claims(s) with clear reasons and relevant evidence, using credible sources and demonstrating an understanding

of the topic or text, (c) using words, phrases, and clauses to clarify the relationships among claim(s) and reasons, (d) establishing and maintaining a formal style, and (e) providing a concluding statement or section that follows from the presented argument (Common Core State Standards, NGACBP & CCSSO, 2010).

While grade 11-12 students are expected to adhere to the aforementioned content standards, they are additionally tasked with extending and refining their ability to introduce precise and knowledgeable claim(s), differentiate the claim(s) from alternate or opposing claims, and devise an organization that logically sequences claim(s), counterclaims, reasons, and evidence. Moreover, they are required to clarify the relationships between claim(s) and reasons, between reasons and evidence, and between claim(s) and counterclaims. A brief examination of these standards reveals that Michigan's standards (and by proxy, the more broadly adopted CCSS for ELA) aim to develop heightened proficiency in persuasive writing, guiding students toward nuanced argumentation, evidence-based reasoning, and effective communication of ideas.

Learning to write persuasive essays offers numerous personal and psychological advantages. These skills are deemed essential for future academic and career endeavors. Firstly, persuasive writing fosters critical thinking as students are prompted to deeply think about and provide rationales for various issues and advocate for their beliefs (Giri & Paily, 2020; Susilawati et al., 2019). Secondly, it hones students' ability to effectively and logically articulate and organize ideas by leveraging their skills in researching, analyzing, and applying prior knowledge to support their main arguments (Kim et al., 2021; Rubiaee et al., 2019). Thirdly, it encourages students to consider diverse perspectives, fostering empathy and an understanding of the concerns and values of others through perspective-taking approaches (Cho et al., 2021; Hung & Wyer, 2014). Additionally, engaging in persuasive writing can enhance students' confidence in

their communication abilities (McElligott, 2014). Moreover, these persuasive, reasoning, and analytical skills acquired through persuasive writing during K-12 schooling can be applied across various contexts beyond secondary education, contributing to success in different aspects of life (Streibel, 2014).

2.3.3 Performance

Composing persuasive papers is universally acknowledged as a challenging mode of communication for K-12 students. Consistent findings from national and international assessments indicate that students face greater challenges in persuasive writing tasks compared to informative or narrative writing tasks (Applebee et al., 1994; Mo & Troia, 2017b). Ferretti and Graham (2019) note the "gradual development" of written persuasion. Nippold and Ward-Lonergan (2010) also characterize persuasive writing as a "demanding communication task that requires sophisticated cognitive and linguistic abilities" (p. 238). Typically, students encounter various challenges when tasked with writing persuasive essays, primarily owing to the complex nature of this form of discourse.

One primary factor contributing to these challenges is the limited exposure students have to persuasion until the secondary grades. Instructional focus has traditionally been placed on narrative writing up to that point (Applebee, 1986). The introduction of persuasive writing typically begins in middle school, where students are tasked with constructing persuasive pieces supported by evidence to influence a target audience. The experience with persuasive essays deviates from everyday language or oral arguments (R. Andrews et al., 2009), making students less familiar with the nuances of persuasive text structure and topics. When composing essays, students must enhance their awareness of the persuasive intent, standpoint on the given issue, context of the argument, and their understanding of the dilemma at hand (Jonassen & Kim,

2010). While children exhibit audience awareness in oral argumentation from a young age, the shift from oral debate to understanding and articulating written arguments presents challenges that are gradually overcome through formal and explicit instruction (Coirier et al., 1999; Stein & Bernas, 1999).

Secondly, writing persuasive essays in secondary school may introduce novel rhetorical and linguistic challenges for students. Effective academic persuasive writing often demands organizing discourse not merely around a sequence of events but "by employing a stepwise argumentative structure to a series of ideas, frequently incorporating later-acquired discourse markers" (e.g., *nevertheless*, *on the one hand*; Uccelli et al., 2013, p.40). Academic persuasive essays also surpass the mere expression of emotions or reactions toward events and necessitate that writers articulate their stances toward specific ideas, such as expressing the degree of certainty about their assertions (e.g., *it might be that*, *certainly*; Berman & Nir-Sagiv, 2004). Additionally, a persuasive text, akin to other expository subgenres (i.e., compare/contrast, cause/effect, problem/solution), poses inherent challenges related to use of diverse and complex sentence structures (Karasinski, 2023; Deng et al., 2022).

Thirdly, composing a persuasive written argument demands a spectrum of higher-order abilities, including perspective-taking, critical thinking, and problem-solving, contingent upon a robust understanding of the subject matter (Toulmin, 2003). The acquisition of these is expected to lay the foundation for the discipline-specific formal argumentation anticipated in subsequent secondary school years (Hillocks, 2002). Interestingly, proficient writers appear to be most cognitively engaged when writing persuasive essays and least engaged when writing other genres such as descriptive essays, suggesting recruitment of more higher-order intellectual abilities (Bouwer et al., 2015). Some persuasive writing tasks also require students to engage in close

reading of a source text, followed by comprehending the essence of key ideas, internally debating opposing ideas, and integrating ideas from source material into their own writing. However, students in middle school are not frequently exposed to reading materials incorporating extensive argumentation (Chambliss, 1995), implying that many students will struggle with this register.

Fourthly, young writers in early middle school may experience difficulties in effectively translating ideas and knowledge to communicate with a specific group or audience (Scardamalia & Bereiter, 1987). In the case of persuasive essays, students are usually required to articulate arguments in preparation for a classroom discussion, where other students are identified as the intended audience. However, given that the teacher often assumes the role of the discussion leader, students might craft argumentative essays with the teacher as the intended audience. This potential shift in audience could contribute to variability observed in students' persuasive writing performance. For example, grade 6 students were found to face more challenges when composing persuasive papers for their teacher compared to when writing for a peer, resulting in increased score variance (Crowhurst & Piche, 1979). Existing scholarly research on writing assessment provides conclusive evidence regarding a relationship between various components of the writing task, such as communicative purpose and intended audience, and the overall quality of writing (e.g., Bouwer et al., 2018).

Fifthly, students might bring inherent biases to the discussion of a given issue. They might hold pre-existing notions about the topic or subject, which poses a challenge in objectively assessing viewpoints and opposing views with regard to the evidence presented rather than relying on preconceived notions (Perkins et al., 1991). Effectively analyzing, evaluating, and generating persuasive texts requires awareness of one's biases, the capacity to dismantle those

biases, questioning the authority of the source material, and discerning biases from within the text—a constellation of activities many students find challenging, especially at the middle and high school levels (Aziz & Said, 2020; Boyle & Hindman, 2015).

2.3.4 Theoretical Frameworks

In the domain of persuasive writing, some theoretical frameworks or perspectives serve as pillars that influence scholarly discourse and provide a structured foundation for framing research inquiries. These frameworks/perspectives offer a systematic approach to mapping out existing research and streamline the synthesis and assessment of literature pertinent to the research questions and design outlined in Chapter Three. Moreover, they facilitate identifying gaps and limitations in prior work, providing insights into how these contributions align with the broader landscape of writing scholarship. In this section, three prominent theories underpinning the inquiry in this area are introduced and elucidated, with relevance to diverse aspects of persuasive writing and writing instruction.

2.3.4.1 Ethos, Logos, Pathos

Ethos, logos, and pathos represent the three rudimentary components of persuasive rhetoric introduced by the ancient Greek philosopher Aristotle. To achieve rhetorical efficacy and persuasive success, an author must engage the audience through a nuanced crafting of their argument, strategically considering the means by which audience agreement can be attained (Wachsmuth et al., 2018). Aristotle bestowed these modes of engagement with the Greek terms we continue to employ today.

Logos denotes the realm of logic, reason, and rationality. An author relying on logos employs logic, thoughtful structure, and objective evidence to appeal to the audience's intellect. This involves furnishing information amenable to fact-checking and presenting thorough

explanations to substantiate key points. Logical appeals rest upon rational modes of thinking, including elements such as elaboration, cause/effect, deductive reasoning, inductive reasoning, comparison, exemplification, elaboration, and coherent thinking.

Pathos entails authors tapping into the audience's emotions to garner agreement with the author's claim. Authors employing pathetic appeals seek to evoke a spectrum of emotions, including anger, pride, joy, rage, or happiness. Pathos-based rhetorical strategies aim to induce an audience to open up emotionally to the topic, the argument, or the author. By leveraging emotions, an author can exploit the audience's vulnerability, leading them to perceive the argument as compelling. This may involve the use of emotionally laden language, expressive descriptions, and vivid imagery to immerse the reader in a specific emotional mindset.

Ethos appeals include audience values and authorial credibility or character. When authors make ethical appeals, they seek to resonate with the values or ideologies held by the audience. This connection to values can evoke a sense of what is morally right, fostering an alignment with the author's argument. The ethical appeal is intertwined with ethos in the sense of authorial credibility. The trustworthiness of the author is determined by their knowledge, expertise in the subject matter, and personal history and traits. Figure 2-1 illustrates the interconnection among the three persuasive appeals of rhetoric within the context of composing persuasive essays.

Figure 2-1 Rhetorical Triangle (Ramage et al., 2016, p. 55)

The introduction of the Aristotelian rhetorical theory in this context arises from the recognition that incorporating its three appeals in writing is widely perceived as enhancing the effectiveness and persuasiveness of discourse. This observation holds true not only in everyday informal writing, e.g., on social media (Marcotte & Stokowski, 2021; Nelzén, 2018) but also in formal writing genres like legal discourse (McCormack, 2014; Smith, 2014). It extends into educational settings where learners are instructed to manifest writing behaviors and use techniques associated with the three appeals (FitzPatrick & McKeown, 2021; Khairuddin et al., 2021; Mohamad et al., 2022), all aligned with the overarching objective of persuading their intended readers.

Within educational settings, writers seeking to appeal to logos often demonstrate proficiency in presenting factual information, statistical data, and various forms of logical evidence to substantiate their arguments. The efficacy of these strategies can be assessed through a quantitative and sorted framework (e.g., Toulmin's Model, discussed later in this section) which delineates specific elements essential for contributing to the provision of evidence and qualifications. To appeal to pathos, writers strategically deploy language choices related to

intonation, stylistic components, and emotionally charged vocabulary. This strategic use of such linguistic tools can be methodically assessed through sentiment analysis of persuasive essays, aiming to elicit an emotional response from the audience. Establishing ethos, on the other hand, involves the writer positioning themselves as an authority on the subject matter. This is often achieved by expressing sufficient content knowledge and a commitment to good character and intentions. Cultivating ethos through discourse relies on the use of rhetorical language to mediate between the speaker and the audience, thereby projecting character traits that evince credibility (Rideout, 2016).

Durst et al. (1990) conducted an analysis of persuasive essays composed by high school juniors and seniors with the aim of identifying the specific rhetorical and linguistic features that contributed to raters' holistic judgements about the overall quality of the essays. The study incorporated three primary persuasive appeals—ethos, logos, and pathos—grounded in Aristotle's rhetorical theory of persuasion. Their analytic system, a variation of the one developed by Connor & Lauer (1985), employed 23 persuasive appeals to identify features contributing to text persuasiveness. Using a comparable four-point scale to measure students' utilization of the persuasive appeals (refer to Appendix 5: Persuasive Appeals Rubric in their study for detailed descriptions), Durst and colleagues found notable relationships in correlation analysis with holistic scores. Logical appeals demonstrated a substantial correlation with $r = 0.73$, while pathetic appeals exhibited a correlation of $r = 0.38$, and ethical appeals had a correlation of $r = 0.27$, all statistically significant. In a stepwise regression analysis, the use of logical appeals accounted for approximately 53% of the variance in holistic scores. Students demonstrated more frequent and effective use of logical appeals compared to ethical or pathetic appeals. This discrepancy in appeal usage may be attributed, in part, to teaching practices. The

high school writing curriculum placed greater emphasis on the use of logical appeals, with significant instruction on the development of robust reasons and convincing arguments. In contrast, secondary language arts instruction allocated relatively less attention to engaging audience attitudes, values, and emotions, or presenting a caring and knowledgeable image. Overall, this study offers a detailed exploration of rhetorical elements to understand the nature of persuasive writing among high school students.

2.3.4.2 Toulmin's Model of Argumentation

The Toulmin model of argumentation, originally devised to explicate the macrostructure of argumentative essays and nowadays widely adopted as a foundational framework for scientific argumentation within the field of postsecondary education, holds a prominent place in this section. Its inclusion as one of the significant models is justified by its relevance in examining students' persuasive writing within prior scholarly works (Aziz & Said, 2020; Junaidah Januin, 2021; Liu & Wan, 2020), and its significance in shaping research study design. Importantly, this theoretical model played a pivotal role in the annotation of persuasive elements within the PERSUADE 2.0 Corpus (see Crossley et al., 2022, for details), which is the database used in this secondary data analysis study.

Toulmin (1958) formulated his method of argumentation based on a model comprising three essential components integral to persuasive essays. According to this model, an individual (1) puts forth a *claim* (or main argument), then (2) provides *grounds* to substantiate that claim, and (3) backs the grounds with a *warrant* (Karbach, 1987, p.81). Toulmin's model also further incorporates three additional elements: (4) providing supplementary *backing* to support the warrant, (5) acknowledging an alternative viewpoint through *rebuttal*, and (6) introducing *qualifiers* to indicate that a claim may not universally hold true in all circumstances. These

additional elements may be incorporated as necessary. Figure 2-2 illustrates an example of how these elements can be depicted in a persuasive essay.

Backing
Research links phone use in schools to declining academic performance.

Warrant
Restricting phones in schools creates a more conducive learning environment.

Mobile phones should not be allowed in schools.

Grounds
Phones in class lead to reduced attention and lower academic performance.

Qualifier
Acknowledging technology's value, clear policies are needed to ensure responsible use in schools.

Claim

Rebuttal
While some argue for phones as educational tools, challenges in effective integration remain.

Figure 2-2 Toulmin's Model of Argumentation with Examples

In the persuasive genre, developing writers often lack text structure knowledge to differentiate a persuasive schema from more prevalent text structures, which may result in integrating non-argumentative components when responding to persuasive prompts (Crowhurst, 1990). Even when children demonstrate awareness of the structural elements inherent in the argumentative/persuasive genre, they may still struggle to effectively utilize all the essential elements to construct a coherent essay (Wingate, 2012). Moreover, the organization of these structural components to establish a logical flow within the paper proves challenging for school-age writers (Calfee & and Chambliss, 1987).

Wolfe et al. (2009) discovered that both persuasiveness and perceived quality of persuasion could be enhanced by presenting and rebutting counterarguments. Although Toulmin's framework is significant for emphasizing the importance of addressing alternative viewpoints when making claims, researchers face challenges in consistently applying the model. This difficulty arises partially because students' arguments often span multiple elements. For example, claims may be implicit in persuasive discourse and require deduction. Ambiguity in identifying data, warrants, and backing can lead to coding difficulties and reduced reliability (Simon, 2008). Nevertheless, Toulmin's model of argumentation remains a prominent framework for classifying text elements in written persuasion (see Aziz & Ahmad, 2017; Aziz & Said, 2020; Magalhães, 2020) and understanding students' justifications as the basis of their claims and conclusions by analyzing the interrelations between surface structure and substance (Stapleton & Wu, 2015).

Zimmerbaum (2014) recognized the necessity for a paradigm shift in ELA classroom instruction at the secondary school level. In response, an instructional framework was developed, employing a critical questions strategy and the Toulmin model of argument. This framework aimed to assist nine eighth-grade participants in developing the essential skills required for constructing logical, reasoned arguments within persuasive writing. The analysis of pretest and posttest results indicated a notable increase of 1.22 points (on a 5-point scale) in the average persuasive writing scores of the nine secondary students. This enhancement was attributed to guided discussions and practice sessions focused on gathering evidence, scrutinizing evidence for claim development, employing critical questions for claim testing and revision (e.g., Where does the topic occur in the text? Whose actions relate to the topic? When do these actions occur?; refer to Appendix O in Zimmerbaum, 2014 for details), and utilizing the Toulmin model as a

guiding structure to organize their thoughts about a claim. These outcomes suggest that the incorporation of the Toulmin model of argumentation can play a pivotal role in aiding secondary educators as they nurture the development of argumentative and persuasive skills among their students. Persuasive essays rely on the depth of students' prior knowledge and the research they can access and comprehend, as claims and evidence must come from these sources. While the Toulmin model helps assess students' ability to analyze and apply information—particularly in evaluating claims, evidence, and rebuttals—it does not fully address how students activate prior knowledge or conduct research. The effectiveness of this approach depends heavily on how teachers use it to engage students with data and guide them in developing stronger, evidence-based arguments.

At the secondary level, persuasive and argumentative essays often find prominence within disciplinary content area classrooms, particularly in subjects like social science and natural science, where students are instructed to articulate, elaborate, reflect upon, and synthesize the laws, theories, principles, and concepts introduced to them (Sampson et al., 2013). Recognizing this, some researchers have extended the application of Toulmin's model to other disciplinary classes in secondary schools. For instance, Giri and Paily (2020) conducted a study focused on exploring the efficacy of Toulmin's argument structure within the Think-Read-Group-Share-Reflect (TRGSR) scientific argumentation strategy for enhancing critical thinking among secondary students. This quasi-experimental study employed a pretest-posttest control group design, involving 50 twelfth-grade students in total. The experimental group received instruction on Toulmin's argumentation, while the control group followed the traditional teaching approach. Diverging from previous studies, the primary outcome variable in this research aimed to discern differences in students' critical thinking abilities (measured via multiple choice or true/false

items tapping inferencing, deduction, assumptions, interpretation, and evaluation on the Watson-Glaser critical thinking appraisal Form S) between the two groups. The results revealed a demonstrable improvement in critical thinking ability among students in the experimental group. The teaching strategy incorporating the Toulmin components proved to be relatively more effective ($F = 83.12$, $p < 0.001$) compared to the control group's traditional teaching approach.

2.3.4.3 Writer(s)-Within-Community Model

The Writer(s)-Within-Community Model (WWC; Graham, 2018) posits that writing and writing instruction are influenced by the communities or contexts in which they occur, as well as by the cognitive abilities and resources of writers and educators involved in writing tasks within these settings. Developed writers typically undergo five major production processes to accomplish writing tasks (Steiss et al., 2024): conceptualization (creating a mental representation of the task), ideation (generating, developing, and conveying new ideas), translation (converting thoughts into language), transcription (spelling, handwriting, and keyboarding), and reconceptualization (revising and editing). In the current study, the primary focus is on the reconceptualization process, which aligns with the secondary purpose of this research to use MMFP to structure formative feedback provided by generative artificial intelligence (GenAI, specifically GPT). We posit that GPT-generated feedback based on derived MMFP models can aid students in refining their drafts by assessing how their writing aligns with ideal genre standards and providing specific steps for improvement. This assumption is grounded in empirical findings from prior studies, which have shown that writers benefit from multiple sources of information regarding revision, including peer review (Wu & Schunn, 2021), self-assessment (C. S. Johnson & Gelfand, 2013), and teachers' corrective feedback (Link et al.,

2022). These approaches have been proven effective in enhancing the quantity and quality of revised drafts and overall writing proficiency.

Furthermore, the importance of reconceptualization is particularly highlighted for secondary students involved in persuasive writing tasks. This cohort of learners is at a critical stage in developing persuasive writing skills, which necessitates sophisticated language use and advanced cognitive abilities (Masrul & Yuliani, 2023). In addition, the variation in writing contexts, such as writing purposes, the perceived value of writing, types of writing tasks, and common writing practices, can all significantly influence students' learning experiences and impact the quality of their written products. Effective writing instructions ideally entails teachers providing frequent and personalized feedback across multiple drafts; however, this process is usually time-consuming and resource intensive. Therefore, well-structured and informative written feedback promptly provided by GenAI is presumed to be more advantageous for secondary students in enhancing their persuasive writing compared to other approaches.

2.3.4.4 Genre Pedagogy

In addition to the two theoretical frameworks for how to persuade/argue (the Aristotle and Toulmin frameworks), it is imperative to delineate an instructional framework in this subsection –genre-based pedagogy–which is informed by the sociocognitive pragmatics-based framework and other theoretical perspectives such as Halliday's systemic functional linguistics (1994). The sociocognitive pragmatics-based framework regards oral and written language use as socioculturally situated cognitive practices (Bazerman & Paradis, 1991; Snow & Uccelli, 2009). This perspective implies that, as writers mature and navigate a broadening array of language-mediated social contexts, they continually acquire new modes of speaking and writing, often shaped by linguistic and cognitive demands (Uccelli et al., 2013). Certain text genres are

considered more linguistically and cognitively demanding than others, and individuals vary significantly in their exposure to different communicative contexts (W. Qin & Uccelli, 2016). As a result, it is anticipated that learners may excel in writing within one genre but encounter challenges in another. Secondary academic writing, representing a crucial developmental stage in education, can thus be conceptualized as integral to the growth of young writers. Secondary writing experiences expose students to various disciplinary content areas and help them acquire a flexible repertoire of language forms and functions across a broadening range of social contexts, thereby preparing them for advancing to post-secondary education (Lavelle et al., 2002).

Educational linguists have elaborated on the effectiveness of genre pedagogy in teaching learners how to use genre-specific linguistic resources to convey content knowledge, engage in interpersonal relationships, and organize texts within the academic genres that adolescents encounter (Christie, 1998; Rose & Martin, 2012). For instance, in the study conducted by Ramos (2015), researchers implemented a genre-based pedagogy instructional approach that integrated components such as building vocabulary knowledge, facilitating understanding and discussion of ideas and concepts, and offering explicit instruction on linguistic resources, projecting authoritative stances, and constructing well-organized texts. The observed improvement in participants' writing from pretest to posttest in this study was attributed to the students' enhanced control over the linguistic resources required to compose academic persuasive essays.

In alignment with the Common Core State Standards (NGACBP & CCSSO, 2010), which mandate educators to equip all K-12 learners with the skills to cultivate academic literacy practices, the adoption of genre pedagogy emerges as a potential avenue to achieve this educational objective. In a study conducted by Ramos (2019), twenty students spanning grades 9 to 12, selected from a U.S. northeastern urban public high school, were subjected to an

instructional intervention known as the "reading to learn" approach rooted in genre pedagogy. This pedagogical approach involved guiding students through the process of deciphering unfamiliar vocabulary, concepts, and idiomatic expressions within persuasive texts. Additionally, the intervention sought to foster the development of background knowledge concerning common arguments related to the subject topic, along with enhancing students' linguistic awareness by directing their focus towards metalanguage elements (such as nominalization and causal verbs) embedded in the reading materials. The results of the study demonstrated a substantial increase in participants' effective utilization of key academic linguistic resources aimed at constructing persuasive discourse, as evidenced by the observed shift in the effective use of the 14 linguistic resources measured (e.g., nominalization, projection, evaluative language, modality) from pretest to posttest. These findings imply that the genre pedagogy-based approach holds promise in facilitating the development of academic literacy practices among adolescents. This implication also guides this study's design, as a broad range of genre-specific language features that characterize persuasive writing will be explored. These features will be thoroughly addressed in the next section.

2.4 WRITING MEASUREMENT

2.4.1 Features

When examining the features of academic writing, educational researchers consistently direct their attention to both the cognitive processes involved in writing and the resultant written product (S. A. Crossley, 2020; M. D. Johnson, 2014; Torrance et al., 2021). Hayes' cognitive model of writing (1996) outlines three principal stages in the writing process: planning (pre-writing), drafting (composing or translating), and reviewing (editing or revising). Proficient writers demonstrate a capacity to monitor their processes and progress throughout the entire

writing endeavor, ultimately producing an acceptable final written product (Graham & Harris, 2013; Graham & Perin, 2007; Laist, 2021).

This study's secondary data analysis specifically emphasizes the attributes of the writing product over writing processes due to: (1) the absence of features related to writing processes in the PERSUADE 2.0 corpus and (2) the belief that features of the writing product can serve as reflections of students' writing processes, aligning with prior work on the effects of writing process instruction on writing outcomes (e.g., Cutler & Graham, 2008; Graham & Sandmel, 2011; G. A. Troia & Olinghouse, 2013). To delve further, during the pre-writing phase, writers who sophisticatedly organize ideas and establish writing goals exhibit a propensity for generating a writing product characterized by clarity, well-structured language, technical terminology, and creative linguistic choices to explicate the subject matter (De La Paz & Graham, 1997; Olinghouse, 2008). Similarly, engagement in the reviewing phases, wherein writers modify their drafts, tends to yield a high-quality writing product with fewer language and mechanical errors and enhanced coherence (De La Paz & Sherman, 2013; Shen & Troia, 2018b).

When evaluating written products across various discourse genres, particularly in persuasive writing that requires multifaceted writing skills, researchers routinely delve into the analysis of microstructural and macrostructural elements (S. Hall-Mills & Apel, 2015; Karasinski, 2023; Richards, 2013). Examining discourse elements at both levels within a text holds considerable promise for capturing the variations in language features across diverse groups in their written expression. The subsequent discussion elaborates on macrostructural and microstructural elements due to their significance for feature extraction and selection in the subsequent data analysis section.

2.4.1.1 Microstructural Elements

The examination of microstructure elements in written products involves various levels of language, including word, sentence, and discourse levels. Puranik and colleagues (2008) identified productivity, accuracy, and complexity as significant dimensions warranting investigation in students' writing. In a study by Troia, Shen, and Brandon (2019), multiple measures of written expression were explored as predictors of narrative writing performance for 362 students in grades 4 through 6. At the word level, the token-type ratio for content lexemes indicated word productivity, while mean textual lexical diversity, mean content word frequency, and mean syllables per word acted as proxies for word complexity. Word accuracy was reflected by the proportion of words absent spelling and/or capitalization errors. Moving to the sentence level, metrics such as the percentage of complex sentences, mean words per sentence, and mean words before the main verb offered insights into syntactic complexity. Indicators of sentence accuracy included the percentage of grammatical sentences and mean punctuation errors in a sentence. Finally, at the discourse level, the total number of words written reflected text productivity (which was redundant with number of sentences), while the incidence of connectives and narrativity score served as indicators of discourse complexity.

These writing metrics have contributed to a burgeoning body of literature advocating for the adoption of a levels of language framework in the assessment of writing quality. The utilization of a combination of indices encompassing word, sentence, and discourse levels has demonstrated effectiveness in gauging distinctions among written products and the individuals accountable for their creation (see Carvalhais et al., 2021; M. Kim & Crossley, 2018; Sarmiento et al., n.d.; G. A. Troia et al., 2019; Wang & Troia, 2023b). These findings are consistent with existing writing research, highlighting robust correlations between factors such as spelling

accuracy, sentence grammaticality, and text length with writing quality. These correlations effectively differentiate between mature and immature writers while also showcasing variations across different grade levels. The subsequent subsections expound upon the empirical findings pertaining to students' performance at the word, sentence, and discourse levels in their written expression, while also examining variations in writing performance among students with diverse cultural and language backgrounds.

<u>Word level</u>

Hayes' (1996) cognitive model of writing emphasizes the cognitive and linguistic resources essential for producing high-quality text. Long-term memory is a pivotal cognitive resource through which writers translate and interpret ideas, experiences, and sensory images into linguistic form (Olinghouse & Wilson, 2013). Adept use of vocabulary can facilitate this complex process. Another theoretical framework validating the important role of the lexicon in text is Scardamalia & Bereiter's (1987) knowledge-telling model, which further elucidates the links between vocabulary and long-term memory. For instance, vocabulary can convey content knowledge, given that many topics necessitate specialized terminology (Harmon et al., 2005). Vocabulary is also implicated in discourse knowledge, with the assumption that it can discriminate the characteristics of different genres of text (Biber, 1989; Hasan, 2014).

The acquisition of vocabulary demands substantial cognitive effort and knowledge resources, particularly during middle and high school (Elleman et al., 2019; Groves, 2016). However, vocabulary acquisition is further complicated by the diverse nature of words, some of which may have different linguistic and grammatical forms, such as polysemous words (Bowers, 2011). Students may also face challenges when confronted with idiomatic expressions, making it difficult for them to choose the appropriate meaning of words (Rohmatillah, 2014). These

challenges in vocabulary learning, when coupled with insufficient and ineffective vocabulary teaching strategies and interactions, can give rise to more serious problems. These may include difficulties in achieving writing proficiency, language-based problems, elevated dropout rates, and a substantial word gap compared to peers (Armstrong et al., 2018; Duff & Brydon, 2020; Graham et al., 2016).

Apart from recognizing the cognitive challenges of vocabulary acquisition often experienced by students, Olinghouse and Wilson (2013) conducted a quantitative study to describe the surface-level characteristics of successful writing concerning vocabulary types. This research serves as an empirical guide for understanding aspects vocabulary instruction should entail, guiding educators on where to place emphasis when teaching students to enhance their vocabulary and evaluate their writing performance. The study investigated the relationship between the targeted vocabulary constructs and text quality across three genres (narrative, informational, opinion). The findings disclosed variations in the type of vocabulary used by fifth-grade students across these genres. High-quality narrative writing showcased greater vocabulary diversity and maturity but less elaboration and fewer academic words. Conversely, strong persuasive compositions featured higher vocabulary diversity and a more formal register. High-quality informative texts exhibited a greater number of content words, increased elaboration, and higher maturity levels. These findings suggest that word selection in writing appears to be influenced by genre, with somewhat different vocabulary constructs contributing to writing quality depending on the genre being scrutinized.

In Sarmiento et al.'s (2024) study, researchers extended Olinghouse and Wilson's (2013) investigation by shifting the focus to vocabulary constructs that represent the academic register in grades 5 and 8. The online automated text analysis tool Coh-Metrix was employed to extract

vocabulary features, including mean lexical diversity, mean number of words per sentence, left embeddedness, total number of words, and incidence of causal and adversative connectives. These indicators served to assess students' use of academic language via lexical diversity and density. The study also considered long words (words with 7 or more letters) and the frequency of academic words within students' essays to represent their use of complex and academic language. Results indicated that diverse vocabulary predicted writing quality in grade 5 but not in grade 8, but the number of long words was related to both writing quantity and quality at both grades. The researchers suggested that educators can identify characteristics of these long words for targeted instruction to enhance students' academic success. For instance, proper nouns, nominalizations, gerunds, hyphenated words, and words with affixations are believed to significantly improve both the quality and quantity of students' academic written language.

To sum up, the lexicon assumes a pivotal role in both constructing and interpreting meaningful text (Engber, 1995, p.141). As academic tasks become increasingly complex throughout middle and high school, writers are expected to deploy academic vocabulary effectively to achieve success in their scholarly endeavors (W. Nagy & Townsend, 2012). Hence, students' vocabulary knowledge and lexical competence hold a fundamental position in quality text production (Maamuujav, 2021). Numerous studies have indicated significant and positive associations of various vocabulary-relevant constructs, such as vocabulary size (LAUFER & NATION, 1995; Tömen, 2016), lexical diversity (Gómez Vera et al., 2016; Sadeghi & Dilmaghani, 2013), use of sophisticated words (Harb, 2018), academic words (Y.-S. Kim et al., 2013; Sarmiento et al., 2024), and word spelling accuracy (S. Andrews et al., 2020; G. A. Troia et al., 2019) with text quality.

Prior scholarly research has yielded valuable insights into the types of words deemed suitable for instruction in general writing classrooms. However, students from diverse populations, especially those facing cognitive and psychological challenges, require a nuanced approach to vocabulary learning. For instance, children at risk for learning disabilities often manifest neurologically-based developmental delays in perception, short-term and working memory, self-regulation, phonological awareness, and orthographic knowledge (Breznitz, 1997; Wolf, 1997). These deficiencies can pose challenges in learning vocabulary (e.g., technical terminology), making it difficult to comprehend the meanings, relationships, and conceptual interpretations of important terms (Bryant et al., 2003; Kame'enui & Baumann, 2012). Prior studies have indicated that children with language-based learning disabilities exhibit lower productivity (Mackie & Dockrell, 2004; Scott & Windsor, 2000), reduced lexical diversity (Fey et al., 2004), and less grammatical accuracy when using vocabulary with inflectional endings (Dockrell et al., 2007; Gillam & Johnston, 1992) compared to their typically achieving peers when tasked with composing essays. Similarly, non-native English speakers may confront challenges in vocabulary acquisition, influenced by reasons often rooted in cultural factors. Understanding a word involves familiarity with its literal meaning, varied connotations, syntactic constructions it entails, morphological options it presents, and a rich array of semantic associates such as synonyms and antonyms (W. E. Nagy & Scott, 2000). Dobbs and Kearns (2016) found that English language learners with limited English proficiency were less likely to use and attempt words than their native speaker counterparts when analyzing their language arts standardized testing (i.e., Massachusetts Comprehensive Assessment System) results.

Sentence level

Syntax refers to the set of rules governing the arrangement of words into larger meaningful units, such as phrases, clauses, and sentences (Kamhi & Catts, 1999). Understanding, activating, and applying syntactic knowledge has been linked to enhanced writing quality in students (Cain, 2007). This correlation is apparent throughout the school years, starting as early as grade 4 when students incorporate more intricate syntactic structures specific to genres in their written compositions (Furey et al., 2017). This developmental progression tends to plateau around grade 7 or 8 (A. J. Truckenmiller & Petscher, 2020). The coherence of writing relies significantly on the adept use of precise and varied grammatical sentence constructions (Witte & Faigley, 1981). Hence, students' proficiency in comprehending and producing diverse grammatical structures within sentence contexts plays a pivotal role in achieving this (Catts et al., 2006; Cutting & Scarborough, 2006).

Constructing texts with diverse sentence structures introduces complexity to the writing task and provides writers with a means to articulate meanings and intentions precisely (Graham & Harris, 2013). Proficient writers skillfully diversify their sentence structures in accordance with the writing purpose and intended audience. For example, short and simple sentences with minimal variation can facilitate the readability and comprehension of the text for less knowledgeable or sophisticated readers, whereas lengthy and complex sentences can prevent monotony, infuse rhythm, and present ideas in a coherent manner.

Depending on the writing purpose and genre, incorporating syntactically sophisticated sentences enables the expression of complex ideas within a singular structure and elucidates explicit relationships between sentence constituents (Jagaiah, 2017; H. Kim & Ro, 2023; Maamuujav, 2021). For example, in persuasive writing, establishing close causal links between

facts and claims requires manipulating sentence structures to condense intricate information. While complex syntax may present comprehension challenges due to heightened text difficulty, it signifies sophistication in written texts (McNamara et al., 2010). However, the imperative to employ syntactically complex sentences may not be universally applicable across all writing types or registers, given variations in topics, writing purposes, and audiences that impose differing syntactic demands on the writer (Jagaiah et al., 2020). Research suggests that syntactically complex sentences are often associated with the argumentative or persuasive genre, where writers must intricately connect arguments and evidence (Beers & Nagy, 2009).

Considerable research in writing has consistently substantiated that higher quality writing generally incorporates more sophisticated and complex syntactic features (Beers & Nagy, 2011; S. A. Crossley, 2020; Deng et al., 2022; Maamuujav, 2022; McNamara et al., 2014), such as longer sentences, a greater variety of sentential structures, increased clausal subordination, a higher rate of production of dependent clauses, and greater phrasal complexity. An array of writing measures that are used to evaluate syntactic structures include clauses per T-unit, words per clause, mean clause length, mean number of words per sentence, mean number of words prior to the main verb of the main independent clause, etc. However, it is worth noting that some studies argue that simple sentences can also be powerful, and the effectiveness may be contingent on the genre and task types (Beers & Nagy, 2011). For example, Crowhurst (1980) found that persuasive essays with high syntactic complexity scores received significantly higher quality ratings than essays with low syntactic complexity in grades 10 and 12, although no difference was observed in grade 6. Conversely, for narratives, the only significant finding was that, in grade 12, low syntactic complexity was associated with higher quality. This may be

attributed to the fact that students by the end of secondary education can strategically vary sentence structures to express ideas and captivate their audience.

Students with low proficiency in writing and limited writing knowledge, particularly those with learning disabilities and language impairments, often encounter challenges when attempting to craft well-constructed and sophisticated sentences (Saddler & Preschern, 2007). Typically, these struggling writers generate sentences that are less syntactically complex and exhibit consistent grammatical errors (Myklebust, 1980; Puranik et al., 2007). This assertion is supported by previous research; for example, Koutsoftas and Gray (2012) demonstrated that students with language learning disabilities exhibited lower sentence complexity in their written narrative samples compared to their typically achieving peers. However, the two groups did not show a significant difference in sentence complexity measures in expository writing, suggesting that expository writing may necessitate writers to produce more complex sentences than narrative writing.

<u>Discourse level</u>

Analyzing language samples at the discourse level is another vital aspect of assessing students' writing quality. It is important to note that the discourse-level features in this secondary data analysis are related to surface-level language features that were validated in a three-level language framework using confirmatory factor analysis (Wilson et al., 2017) rather than the discourse elements that focus on macrostructural elements to construct the entire text (which will be discussed in the next section), despite that some studies may mix these two aspects together.

Troia, Shen, and Brandon (2019) conducted an analysis of narrative writing samples from 362 students across grades 4 through 6. The researchers identified discourse-level variables, including text productivity (total words written in the entire essay), text complexity (incidence of

connectives), narrativity, and writing process use (evidence of planning, drafting, revising, or a combination thereof), which were significantly associated with writing quality, explaining 15.5% of the variance in quality. It was anticipated that discourse-level features in essay writing contribute to macro-level structures such as organization, style, coherence in content, and other global language conventions that play a substantial role in shaping the overall appearance of the essay.

2.4.1.2 Macrostructural Elements

Macrostructure analysis primarily occurs at the discourse level (Scott, 2009). Macrostructure is defined as the "gist," the "topic of the text," the "summary," or the "overall notion of what the text is about" (Witte & Faigley, 1981). Unlike microstructure dimensions of compositions, which usually involve small-unit and surface-level changes that do not greatly impact the overall meaning of a text, macrostructural elements demand changes at the rhetorical level of writing (S. S. Hall-Mills, 2009; Karasinski, 2023). These changes contribute to shaping global structure and cohesion and can serve as reflections of writers' techniques and abilities. For novice writers, understanding the strategic approach to composing in macrostructural elements can be a challenging endeavor.

Existing literature underscores the significance of discourse knowledge that can largely impact the macrostructural levels of text, which varies based on the assigned topic or type of writing task, not only in early writing development but also in higher grades (Olinghouse et al., 2015; Olinghouse & Graham, 2009; Scardamalia & Paris, 1985). Bereiter and Scardamalia (1987) proposed the knowledge-telling model, highlighting the significance of discourse knowledge in the writing development of young students. According to this model, a child develops a mental representation of the writing task by defining the topic and utilizing their

discourse knowledge to determine the type of text to be produced. As the child writes, this mental framework directs the retrieval of relevant information from long-term memory. The retrieved information is then evaluated for its alignment with the topic and text type, with discourse knowledge playing a role in this assessment. If the content is deemed suitable, it is transcribed into the written text (Olinghouse & Graham, 2009, pp. 39-41). The resulting text then prompts a further search of long-term memory. Essentially, when employing the knowledge-telling approach to writing, young authors likely draw on their understanding of what makes a compelling story and effective writing, alongside relevant content knowledge, to shape the writing task. They use this mental representation to seek out appropriate content from long-term memory. For instance, in Olinghouse and Graham (2009)'s study, an examination was conducted to ascertain whether five types of discourse knowledge—substantive, procedural, motivational, basic story elements, and irrelevant—made a significant and unique contribution to predicting second- and fourth-grade students' story writing performance. The hierarchical regression analyses revealed that the collective contribution of the five types of discourse knowledge significantly predicted overall story writing quality, quantity, and lexical diversity. In the subsequent subsections, some prominent macrostructural elements are discussed through empirical findings.

Effectiveness of text structure

Toulmin's framework of argumentation has predominantly focused on the field-invariant features of an argument by highlighting its six structural elements. However, there exists limited understanding of the effectiveness of arguments constructed by students even when they have adhered to the basic argument structure in their writing. For example, Clark and Sampson (2007) found that students' arguments frequently included incorrect scientific concepts, despite

exhibiting a relatively sophisticated argument structure. This discovery aligns with much other scholarly research that emphasizes the structure of arguments without delving into the content of the argumentation (D. Liu & Wan, 2020; P. Stapleton & Wu, 2015; Zhang, 2018). In response to this gap, scholars in the field of formal reasoning and argumentation have proposed three criteria–acceptability, relevance, and sufficiency/adequacy–to assess the soundness of arguments (Bickenbach & Davies, 1996). Acceptability pertains to premises that are reasonable to accept as true; relevance is a prerequisite for something serving as evidence for the conclusion; and sufficiency/adequacy implies that all premises, considered collectively, should provide enough support to justify belief in the conclusion (see Crossley et al., 2023, for more details). Various schemes illustrating different degrees of detail for assessing the quality of reasoning have also been proposed (Erduran et al., 2004; Means & Voss, 1996).

These criteria provide insightful implications for distinguishing between effective and ineffective persuasive writing. Firstly, it is essential to present ideas in a manner that effectively supports viewpoints, acknowledges alternate perspectives, and illustrates the limited persuasiveness of alternate arguments through comparison to the main arguments. Secondly, the quality of reasoning in arguments should be relevant, accurate, and structurally logical. In other words, both surface structure and substantive content need consideration when evaluating the overall quality of a persuasive or argumentative essay. Sandoval and Millwood (2005) provided insight into the coordination and quality of claims and evidence. Proficient use of structural elements in written argumentation reflects students' tacit understanding of the meaning of these elements and how they support specific claims. Therefore, the orchestration of claims and evidence serves as a cognitive skill signifying an understanding of the data or concepts that underlie claims. More importantly, it is part of a broader social practice employed to persuade an

audience. Students' argumentation and explanations thus mirror their ideas about what makes an argument persuasive to their intended audience.

Tone

Tone, in the realm of writing, refers to the use of language tailored to the audience by incorporating appropriate seriousness and markers of politeness and degree of certainty (Kilgannon, 2022). The CCSS for ELA encompass language standards that emphasize "effective choices for meaning or style" (CCSS, 2010; Aull, 2015). Writers are tasked with maintaining an appropriate tone that considers the potential audience to mitigate resistance. In persuasive writing, the tone serves to convince the audience to align with the author's perspective, taking on an assertive, passionate, or even aggressive quality (Hinkel, 2003; Ho & Li, 2018). The goals of the audience influence various aspects of writing tone, prompting the use of respectful, formal language, markers of politeness, and qualifications. For instance, Midgette, Haria, and MacArthur (2008) present criteria for evaluating the tone of persuasive essays written by fifth- and eighth-grade students, illustrating how the use of respectful, formal language and markers of politeness significantly influences essay coherence and audience engagement. Notably, their study reveals that girls scored higher than boys in persuasive tone, demonstrating greater evidence of adapting language to be acceptable to the reader, resulting in a tone that is more respectful and less prone to rudeness or expressions of anger.

Discourse markers

In persuasive writing, students are expected to adhere to specific language conventions and employ organizational markers (Dobbs, 2014; Uccelli et al., 2013) that are precise and suitable for the communicative context of discussing issues and arguments, albeit not overly specialized to be heavily discipline-specific (Bailey, 2007). Even very young students have

demonstrated attempts at utilizing complex language with discourse markers when engaging in persuasive writing (Nippold et al., 2005).

While limited studies have specifically analyzed proficiency in linguistic conventions marking textual organization within persuasive writing, an array of factors has been identified as contributors to higher quality writing, such as the use of cohesive devices (Crowhurst, 1980), rhetorical devices (Connor, 2004), and hedges and (Alward et al., 2012; Durik et al., 2008). In a study focusing on the academic persuasive writing of high school students, Uccelli, Dobbs, and Scott (2013) found that organizational markers indicating sequences of ideas and certain markers of stance, linguistic indicators conveying attitudes toward propositions, contributed to overall writing quality beyond the impact of lexical and syntactic variables. This current study aims to extend these findings and address a research gap by closely examining the academic language resources that middle school students employ in their persuasive writing.

2.4.2 Issues and Concerns

The assessment of writing has persistently presented challenges for educators and researchers (Fernsten & Reda, 2011; Huot, 1990; Slomp, 2012; Warschauer & Grimes, 2008), with significant concerns revolving around the complex nature of writing, the constrained size of educational datasets, issues related to scoring validity, and the absence of direct implications for instructional practices. Specifically, the growing prevalence of automated essay scoring techniques over the past decade, which compels a shift towards reliance on large-scale datasets and machine-driven algorithms and a departure from the traditional reliance on human raters, has sparked skepticism regarding the robustness and effectiveness of contemporary approaches to writing assessment (Deane, 2013; Hannah et al., 2023; Wang & Troia, 2023b).

The primary cluster of concerns is derived from the fact that writing, by its nature, is a complex, recursive, and sociocognitive activity. This complexity makes it challenging to capture the intended purpose of writing and may also indicate a lack of effective writing constructs involved in measurement to comprehensively assess writing. Deane (2012) proposed a way forward toward a sociocognitive approach to writing analysis. One avenue involves refining the restricted constructs currently involved in automated assessment systems to replicate human scoring in high-stakes contexts, while another pathway suggests that a richer conception of writing should be drawn from various sources that consider the social and cognitive dimensions of writing, as suggested by Flower (1994) and Hayes (2012). The constrained viewpoints in writing assessment may also be attributed to the inclination of researchers to predominately concentrate on scrutinizing microstructural elements, which are frequently identified as responsive to developmental changes across all K-12 grade levels (R. Berman & Verhoeven, 2002) and are deemed more straightforward to identify and extract, in contrast to macrostructural features that represent the abstract portrayal of the global meaning structure (Sanders & Schilperoord, 2006) and that prove to be a more challenging aspect to capture.

To capture the complexity of writing, a variety of scoring rubrics such as primary trait, analytic, and holistic scoring rubrics are considered for assessing writing quality in distinct ways. However, ongoing debates persist regarding the preferred rubric choice, with challenges emerging from each method. For instance, analytic scoring, associated with specific rhetorical situations and distinct dimensions, encounters difficulties as these dimensions often interrelate rather than being clearly distinguishable (Huot, 1990; Lloyd-Jones, 1977). This interdependence poses validity concerns for raters and may necessitate a time-consuming and costly scoring process (G. A. Troia et al., 2019). Trait scoring rubrics typically employ dichotomous criteria to

determine the presence or absence of construct-relevant ideas in student responses (Kaldaras & Haudek, 2022). Nevertheless, this scoring schema may give rise to challenges associated with imbalanced data (Wang et al., 2024). Conversely, holistic scoring utilizes multi-leveled coding schemes with the objective of delivering a singular, overall judgment of students' writing quality (Jescovitch et al., 2021). Yet, this approach may confront issues related to data granularity, especially concerning polar scoring levels, where students' scores often cluster around the middle levels. To conclude, the selection of rubrics often leads to controversy, with insufficient evidence supporting the universal superiority of one scheme over another in human coding (Tomas et al., 2019). The choice of a scoring method is highly contingent upon the specific writing constructs being assessed and the intended purpose of the overall judgement, yet challenges persist in terms of scoring validity and reliability regardless of the chosen scheme.

Another cluster of concerns centers around the perceived absence of a clearly articulated framework for recognizing the interconnections between assessment results and their implications for instruction and broader assessment systems. Factors such as types/genres (e.g., integrated vs. independent, narrative vs. non-narrative) and purposes of writing assessment tasks (e.g., formative vs. summative), along with the location/situation of how the writing activities occur (in schools vs. outside-of-school, naturalistic vs. guided), contribute to the diversity and complexity of writing assessment. Educators proficient in writing assessment understand the purpose of the assessment, the appropriateness of instruments and conditions, and the interpretation of assessment data (Inbar-Lourie, 2008). Educators can also play a crucial role in bridging the research-practice gaps in writing assessment and shaping assessments aligned with realistic classroom contexts. GenAI offers educators a powerful tool for developing "mentor texts" that align closely with specific rubrics. This functionality enables educators to illustrate

various measurable aspects of writing at both micro- and macrostructural levels in a concrete and accessible manner. However, translating assessment data into actionable insights for teachers can be challenging. Research has shown that assessment data is intertwined with the thoughts and actions of educators, particularly as they model their reasoning processes for students (Ball, 2000; Schunk & Zimmerman, 2007). Hence, assessment data can be conceptualized as a mediator of teaching performance. In the discussion section, this topic will be further explored, focusing on how different types of assessment data can prompt educators to customize their instructional practices effectively to address specific writing features.

Contemporary writing assessments often prioritize enhancing modeled performance on scores rather than providing constructive feedback, which is a significant reason why practitioners may perceive writing assessment as less informative due to inadequate feedback to guide students' improvement (Wilson et al., 2021). Additionally, some assessments encounter ethical concerns and measurement bias as they may, intentionally or unintentionally, overlook variations related to students' language background, presentation of presumed racial/ethnic identity, and beliefs or conventions specific to a given cultural context (Fan et al., 2019), particularly among underrepresented student groups.

In addition to the aforementioned general measurement issues in writing, the evaluation of persuasive essays presents some specific challenges. A critical concern lies in determining the validity of persuasive discourse structures, including the assessment of evidence strength, logical reasoning, and emotional appeal. Raters often apply distinct criteria in evaluating what qualifies as a compelling persuasive element (Christian & Iryna, 2014). Some may utilize a lexicon of surface cue words and phrases to delineate arguments at the sentence level, aiding in the evaluation of individual argument content and enriching the feature set for precise score

determination (Burstein et al., 1998). However, this approach has been heavily criticized, asserting the necessity for a more fine-grained analysis of arguments that includes argument components and their interrelations. Another interesting facet of persuasive writing assessment is its context-dependence, where the effectiveness of persuasive strategies varies based on the intended audience. Research has identified that secondary students tend to employ logical appeals with more persuasive power than pathetic or ethical appeals (DURST et al., 1990; Shermis et al., 2013).

2.5 RATIONALE FOR THE STUDY

In this section, the rationale for the study is presented alongside identified research gaps and insights gained from the literature review. Additionally, the educational significance that the study aims to achieve is discussed.

2.5.1 Research Gaps

Given the documented significance of persuasive writing within secondary education for a diverse student population (Beyreli & Konuk, 2018; Boyle & Hindman, 2015; Thomas, 2014) and its implications for preparing students for postsecondary education and beyond (L. Aull, 2015; Garwood & Van Loan, 2019), this study aims to contribute to the expanding body of literature on persuasive writing assessment and inform pedagogical practices in this domain. The study seeks to address several critical gaps in the extant literature. Firstly, there is a dearth of scholarship examining students' writing performance at the middle or secondary school levels, particularly in the realm of persuasive writing. Much of the literacy research on existing diagnostic measurement historically prioritizes reading over writing (see Paul & Clarke, 2016, for a review), and narrative over non-narrative essays (see Graham, Kim et al., 2023, for a review). Additionally, research concentrating on evaluating writing abilities among secondary

school students often prioritizes objectives such as determining course placement, ensuring program effectiveness, or driving curriculum reform (Bishop et al., 2015; Graham et al., 2005; McKeown et al., 2020), rather than identifying students who may be struggling with writing. These scholarly observations suggest that the available literature in evaluating and addressing writing difficulties among secondary students is lacking compared to the extensive literature focused on reading, native writing tasks, and writing assessment for purposes other than problem identification.

Another gap in the extant research can be identified as the predominant focus of existing diagnostic measures on assessing microstructural elements of writing, such as spelling, punctuation, capitalization, and grammar, rather than macrostructural elements such as organization and text content. Specifically, in the realm of persuasive writing, prior studies have primarily concentrated on evaluating the presence or absence of main elements according to Toulmin's model of argument (F. I. B. A. Aziz & Said, 2020; CRAMMOND, 1998; Junaidah Januin, 2021; Sundari & Febriyanti, 2021) without adequately addressing whether these elements are effectively articulated. Researchers have noted instances where students may include main elements but fail to do so in a manner that enhances persuasiveness (P. Stapleton & Wu, 2015). Moreover, existing research tends to neglect the examination of genre-specific features of text, such as appropriate persuasive tones and content richness. This oversight is significant given that writing standards and admissions tests for college students, such as the ACT, SAT, and GRE, place greater emphasis on macrostructural skills than microstructural skills. Consequently, the current diagnostic measures in secondary writing inadequately assess the types of writing skills that hold utmost importance in university-level writing (Richards, 2013).

There is a dearth of research delineating the characteristics of typical and atypical writing in secondary school students. While some studies offer norms for various aspects of typical secondary level writing (R. Andrews et al., 2009; Uccelli et al., 2013), these norms often exclude students who receive special education services. As a result, it becomes challenging to utilize existing literature to discern differences between secondary school students with and without writing difficulties. A differentiated approach to writing remediation is essential for addressing the diverse needs of learners with varying learning profiles. In this context, GenAI has the potential to customize remediation plans tailored to the specific levels of individual students. Current research also neglects to investigate whether students' sociocultural backgrounds, such as English learner status, might affect the association between writing constructs and essay scores.

In terms of measurement domain, much prior research has concentrated on directly examining the raw writing constructs derived from measurement tools and mathematical algorithms. However, this approach may not comprehensively address all writing constructs in a single investigation and may potentially lead to issues such as multicollinearity, overfitting, and unreliability. It is crucial to mitigate these challenges by reducing the multidimensional nature of writing constructs and categorizing them into latent structures that can be validated through statistical testing methods such as structural equation modeling or machine-based clustering algorithms. In this study, this aspect will be addressed by condensing the numerous writing measures into several latent writing constructs related to microstructural elements (word, sentence, discourse) and macrostructural elements (content, tone/style, structure) through exploratory and confirmatory factor analyses. This process will further elucidate the significance of these measures specifically tailored for persuasive writing.

Finally, it is also worth noting that the current zeitgeist of integrating GenAI into schooling has prompted educators to carefully consider its ethical and effective use within educational settings. Given the paramount importance of writing in both learning and assessment processes, and the strong alignment of writing tasks with GenAI capabilities, there is a critical need to make thoughtful and well-informed decisions regarding how GenAI should be leveraged to support the development of students' writing skills (Escalante et al., 2023). In today's context, with the widespread growth of GenAI in various domains, it can be chaotic as GenAI and its subfields (e.g., prompt engineering) typically necessitate substantial resources and often bewilder users due to their inherent flaws and may potentially undermine public trust. Many scholars are endeavoring to explore how to best harness the capabilities of GenAI (Chan & Hu, 2023). The study objectives align with discourse on this topic by aiming to contribute insights on structuring prompts and optimizing the efficacy of GenAI in educational contexts. This endeavor encompasses considerations of ethical usage, effective integration into pedagogical strategies, and optimizing its potential benefits for students' writing development through organizing input prompts based on the derived MMFP.

2.5.2 Educational Significance

The current study holds significant implications for both researchers and educators in advancing the literature on persuasive writing skills among diverse student populations, considering factors such as gender, English language proficiency, receipt of special education services. Furthermore, this study aims to develop an automated feedback system to inform writing performance for diverse student groups. The anticipated findings are poised to contribute in the following key ways.

Firstly, insights into the writing characteristics influencing student writing quality will deepen understanding of secondary school students' writing abilities, particularly concerning common writing activities at this educational level. Additionally, organizing discrete writing measures into multidimensional latent writing constructs using exploratory and confirmatory methods will bolster subsequent model development and applications, such as validating an automated feedback system for delivering differentiated and individualized feedback and instruction within a formative assessment context. It is also important to consider how feedback is delivered through various practices (e.g., process, strategy, vocabulary, and grammar instruction) and pedagogical tools (e.g., graphic organizers, mentor texts). The discussion section will further explore how GenAI-based learning systems could be utilized to promote writing and examine the potential achievements of our students with this support. Establishing evidence of the importance of writing measures is a crucial validation step, as formative writing assessments play a pivotal role in furnishing educators with actionable information about students' current and projected levels of achievement.

Secondly, the study's comparisons between typical secondary school students and those who receive special education services aim to shed light on the similarities and differences between these groups, aiding in the identification of atypical writing patterns at the middle and high school levels. Furthermore, the study's exploration of persuasive writing features linked to overall writing quality and differences in writing performance across groups seeks to guide educators in targeting specific features during intervention to yield significant improvements in overall writing proficiency.

Chapter 3: Methodology

Chapter Three provides a comprehensive overview of the study's objectives and thorough descriptions of the persuasive writing corpus to be used and the analysis methods to be employed to achieve the objectives. The mechanisms used to prepare the corpus and extract data for analysis are described in detail.

3.1 OBJECTIVES AND RESEARCH QUESTIONS

Building on theoretical frameworks and empirical evidence discussed in Chapter Two, this study investigates the complex dynamics of the existence and effectiveness of language use at both microstructural and macrostructural levels, and its significant impact on the persuasive writing performance of secondary students. The study is descriptive in nature and employs secondary data analysis of an extant large-scale dataset—the PERSUADE 2.0 corpus. The study examines the underlying writing constructs within this corpus, which are then employed to develop and fine-tune a large language model–Bidirectional Encoder Representations from Transformers (BERT)–aimed at predicting human-generated holistic scores of the corpus persuasive essays. Additionally, the study addresses the variability in writing constructs and persuasive performance among diverse student populations. The resulting language model is designed to predict the holistic scores of GPT-revised essays based on specific prompts informed by insights from the MMFP (i.e., Microstructural and Macrostructural Features that underpin Persuasive written composition) system. These scores are subsequently evaluated to ascertain any improvement compared to the original student essay scores. The study addresses the following research questions:

1) What textual attributes serve as optimal indicators of persuasive essay quality in secondary school students?

2) To what extent do secondary students with different special needs status (i.e., students with and without an Individualized Education Plan [IEP]) exhibit significant differences in their holistic writing scores across latent writing attributes?

3) Do essays revised by GPT, a Generative AI application, utilizing prompts derived from factor analysis, demonstrate enhanced performance compared to the original essays written by students?

3.2 DATASET

The K-12 CCSS for ELA highlights the cultivation of argumentation skills in writing instruction, emphasizing the need for students to achieve proficiency in using valid reasoning and providing relevant and sufficient evidence to support claims. Additionally, more sophisticated components such as counterarguments and qualifiers are deemed essential in the compositions of proficient writers. However, the 2012 NAEP Writing Report Card reveals that only about 25% of students' argumentative essays were deemed competent. In alignment with the current educational paradigm, which acknowledges the potential of machine-based algorithms and Natural Language Processing (NLP) tools to analyze large-scale corpora and gain nuanced insights into students' writing performance and capabilities, Dr. Scott Crossley and his research team developed and curated the PERSUADE 2.0 (Persuasive Essays for Rating, Selecting, Analyzing, and Understanding Discourse Elements) corpus. This corpus provides educational researchers with the opportunity to conduct quantitative analyses aimed at understanding the connections between the production of argumentative elements, their effectiveness, and quality ratings, with the expectation that such investigations will offer valuable insights informing classroom practices and assessment.

In recent years, approximately 80% of the PERSUADE 2.0 corpus has been made publicly accessible on the Kaggle website[1], specifically for two NLP related competitions as part of the Feedback Prize series. The corpus comprises a collection of ~25,000 argumentative essays composed by students in grades 6 through 12 from various regions across the United States. Baffour, Saxberg, and Crossley (2023) provided a comprehensive summary of the outcomes stemming from the application of large language model solutions within the Feedback Prize competition series. Their analysis highlighted the identification of potential biases present in the distribution of labels representing outcomes, agreement levels, and demographic representation within the corpus. Such insights underscore the ongoing commitment of educational assessment researchers to refine and enhance algorithmic models based on the PERSUADE 2.0 corpus. Of particular emphasis in these endeavors are considerations related to student demographic factors, including but not limited to race/ethnicity, SES, English language proficiency, and participation in special education. This study is driven, in part, by the imperative to explore and mitigate such considerations within algorithmic modeling frameworks. This study conducts a secondary data analysis of the PERSUADE 2.0 database, which contains annotated essays including evaluations of argumentative elements, argumentation effectiveness, and holistic writing quality (ranging from 1 to 6).

3.3 PARTICIPANTS

More than 25,000 persuasive essays (both source text independent and dependent essays) were gathered for this corpus. Within the dataset, a specific subset of 20,823 textual records incorporates detailed demographic information for individual student writers. The primary focus of this study centers on this particular subset of 20,823 essays for secondary analysis, given the

---

[1] https://www.kaggle.com/datasets/nbroad/persaude-corpus-2

critical importance of demographic information in achieving the research objectives. It is important to note that within this subset of 20,823 essays, a total of 2,979 essays, which contain all the necessary features for this investigation, were utilized to conduct factor analysis and multigroup structural equation modeling. The remaining essays were used to validate the large language model for score prediction and to prompt GPT for essay revision.

The demographic data encapsulates variables including gender (categorized as male or female), grade level (categorized as grade 6, 8, 9, 10, 11, or 12), English language learning status (categorized as native English speaker or non-native English speaker), race/ethnicity group (categorized as White, Black/African American, Hispanic/Latino, Asian/Pacific Islander, American Indian/Alaskan Native, or Multiracial/Other), socioeconomic disadvantage status (categorized as economically disadvantaged or not economically disadvantaged; per Crossley, Baffour et al., 2022, economic disadvantage refers to student eligibility for federal assistance through programs such as free or reduced-price school meals, Temporary Assistance for Needy Families [TANF], or Supplemental Nutrition Assistance Programs [SNAP]), and student special education status (categorized as having an IEP or not). The exploration of group differences grounded in these demographic factors aims to provide valuable insights into distinct sub-populations, thereby contributing to a nuanced understanding of the persuasive essays within the dataset.

Table 3-1 presents a pivot table delineating the demographic characteristics of students included in the corpus disaggregated by grade level. Male and female students are nearly evenly distributed, with male students comprising 48.8% ($n = 10,170$) and female students constituting 51.2% ($n = 10,653$) of the cohort. The students are predominantly native speakers of English (89.4%). In terms of racial/ethnic composition, 45.1% identify as White, while the remainder

constitutes various Black, Indigenous, and People of Color (BIPOC) groups. A majority of students, 53.5%, are identified as not economically disadvantaged. Additionally, 12.9% of students were identified as having a disability and received corresponding special education services during the data collection period. In terms of grade distribution, the cohort includes 1,372 sixth graders (6.6%), 9,629 eighth graders (46.2%), 20 ninth graders (0.1%), 6,315 tenth graders (30.3%), 3,083 eleventh graders (14.8%), and 404 twelfth graders (1.9%). Acknowledging the insufficient representation of grade 9 students, which may potentially compromise the reliability of findings, the decision was made to exclude all pertinent data and corresponding information pertaining to grade 9 to mitigate this issue.

Chi-square tests were conducted to compare demographics across different grade levels (see Table 1). The results indicated significant differences in gender, English language learner status, race/ethnicity, socioeconomic status, and special education status among the five grade levels. The null hypotheses were rejected, and we can reasonably infer that the distribution of students across grades in terms of demographics did not adhere to a uniform pattern.

It is important to note that, to address the first research question, which aims to understand the underlying constructs of persuasive writing essays, a subset of 2,977 essays were selected after excluding grade 9 data ($n = 2$ in the subset). This subset includes all the observed language features necessary for conducting factor analysis. No missing data were detected. The remaining subset will be utilized to address the second research question, which focuses on validating the efficacy of providing improved/revised essays. These essays are derived from those written by peers who scored one point higher on the 6-point scale and from GPT-generated essays prompted based on the MMFP results.

Table 3-1 Demographics and Descriptive Statistics for the Full Sample

	Grade 6	Grade 8	Grade 10	Grade 11	Grade 12	Grand Total	Chi-square tests
Gender							
Male	633	4573	3158	1588	203	10170	
Female	739	5056	3157	1495	201	10653	$\chi^2(4) = 23.563$, p < 0.001
Grand Total	1372	9629	6315	3083	404	20823	
ELL Status							
Native English Speakers	1338	9163	5207	2835	24	18584	
Non-native English Speakers	30	455	990	248	380	2106	$\chi^2(4) = 3761.4$, p < 0.001
Grand Total	1372	9629	6274	3083	404	20782	
Race/Ethnicity Group							
White	732	4741	2600	1290	22	9388	
Black/African American	205	1532	1337	538	55	3682	
Hispanic/Latino	332	2230	1701	662	272	5198	
Asian/Pacific Islander	40	722	338	440	50	1591	$\chi^2(4) = 2390$, p < 0.001
American Indian/Alaskan Native	9	55	50	6	0	120	
Multiracial/Other	54	349	289	147	5	844	
Grand Total	1372	9629	6315	3083	404	20823	
SES Status							
Not Economically Disadvantaged	672	5507	2691	2113	121	11116	
Economically Disadvantaged	700	4058	3624	970	283	9643	$\chi^2(4) = 8033$, p < 0.001
Grand Total	1372	9565	6315	3083	404	20759	
SPED Status							
Not Identified with Disability	1299	8799	4955	2730	336	18135	
Identified with Disability	73	830	1360	353	68	2688	$\chi^2(4) = 657.56$, p < 0.001
Grand Total	1372	9629	6315	3083	404	20823	

Note. The Race/Ethnicity variable was recoded into a binary format (i.e., White as 0 and all other categories as 1). This recoding was necessary because the chi-square test results were not valid due to some cells having fewer than 10 cases, which is inadequate for a reliable chi-square analysis.

3.4 CORPUS VARIABLES

3.4.1 Task Types

In addition to participant variables, the PERSUADE 2.0 corpus encompasses several variables pertinent to the writing task. Specifically, one such variable concerns the type of writing assignment. The corpus comprises two distinct sub-corpora, consisting of source-based essays (n = 11,953) and independent essays (n = 8,870). Source-based writing requires students to refer to a provided text, whereas independent writing excludes this requirement. While students may not be required to have specific background knowledge for the independent set, their general understanding of the topic could still influence their writing process.

3.4.2 Prompt Types

Concerning prompt types, the source-based set derives from seven distinct writing prompts and related sources, while the independent set is derived from eight unique writing prompts. In total, there are 15 unique writing prompts, and all prompts and sources are accessible within the PERSUADE 2.0 corpus.

3.4.3 Word Count

Word count information is also available in the corpus. The mean word count for all students' persuasive essays is 409.23 words, with a standard deviation of 235.78. The minimum word count is 146 words, and the maximum is 8,922 words. According to Crossley, Baffour et al.'s (2022) description, the PERSUADE corpus was restricted to essays with a minimum of ~150 words, of which 75% of the words had to correctly spelled American English words. These filters were implemented to ensure adequate coverage of argumentative discourse elements in the texts and to guarantee that the essays contained sufficient recognizable language output for the

development of NLP features to inform algorithm development, as outlined by Crossley & Kyle (2018).

3.4.4 Discourse Elements

An integral aspect of the PERSUADE 2.0 corpus lies in its incorporation of annotations for essential discourse elements within each essay. The process of annotating essays for key argumentative elements is of importance to advancing our understanding of the argumentative strategies employed by student writers. Furthermore, annotations also help discern how these features contribute to successful writing. To facilitate this annotation process, the PERSUADE research team developed an annotation rubric, which addresses three key dimensions: (a) the identification of argumentative elements in essays; (b) the delineation of relations among these elements; and (c) the assessment of the effectiveness of the identified elements.

Every essay in the corpus underwent human annotation to identify argumentative discourse elements and their interrelations. Adopting a double-blind rating methodology ensured rigorous evaluation, with each essay independently reviewed by two expert raters, and any disparities adjudicated by a third expert rater. A comprehensive exposition of the rating process is available in Crossley, Baffour et al. (2022). The annotation rubric, specifically designed for the identification and evaluation of discourse elements prevalent in argumentative writing, drew inspiration from the seminal works of Nussbaum, Kardash, and Graham (2005) and Stapleton and Wu (2015). These annotation schemes represent adapted or simplified versions of the Toulmin argumentative framework (1958). The discourse elements that were annotated for each essay are:

- **Lead** – an introduction begins with a statistic, a quotation, a description, or some other device to grab the reader's attention and point toward the thesis.

- **Position** – an opinion or conclusion on the main question.

- **Claim** – a reason that supports the position.

- **Counterclaim** – a claim that refutes another claim or gives an opposing reason to the position.

- **Rebuttal** – a claim that refutes a counterclaim.

- **Evidence** – ideas or examples that support claims, counterclaims, rebuttals, or the position.

- **Concluding Statement** – a concluding statement that restates the position and claims.

3.4.5 Effectiveness Scores for Discourse Elements

In addition to being labeled as distinct rhetorical and argumentative elements (i.e., discourse elements) by expert raters, the corpus also includes effectiveness scores that gauge the quality of these argumentation elements. Raters underwent comprehensive training using a standardized rubric, where they were tasked with assigning each element a score of effective, adequate, or ineffective. A detailed description accompanied each category, providing an illustrative example of a discourse element fitting that specific rating.

Crossley and his research team annotated a total of 159,228 elements, ensuring at least a 50% overlap. Among these, ~120,000 elements were scored as adequate, ~32,000 as effective, and ~6,000 as ineffective. Two inter-rater reliability (IRR) statistics were computed: exact agreement and a weighted Cohen's kappa. The overall exact agreement for all elements stood at .718, with specific elements ranging between .68 and .81. Notably, exact agreement tended to be higher for source-based dependent essays compared to independent essays. Cohen's kappa of the effectiveness ratings ranged between .17 and .38 for specific elements, demonstrating a comparatively lower level of agreement. Once again, Cohen's kappa exhibited higher values for

dependent essays as opposed to independent essays. The overall Cohen's kappa for effectiveness across all elements was .32, indicating a fair level of agreement. In instances of disparities in effectiveness ratings, a third expert rater served as an adjudicator to render a final decision. For detailed IRR statistics pertaining to PERSUADE 2.0 effectiveness ratings, refer to Crossley, Baffour et al.'s preprint paper.

3.4.6 Holistic Essay Score

The primary focus of this investigation centers on the holistic essay score, which serves as the dependent variable for statistical analyses and for subsequent language model development to predict human-assigned scores accurately. The evaluators responsible for annotating discourse elements for the corpus essays and assessing the effectiveness of each element also holistically scored each essay. In the case of independent essays, raters underwent training utilizing a standardized SAT (i.e., Scholastic Aptitude Test) holistic essay scoring rubric with a 1 to 6 scale; a score of 6 signifies an essay demonstrating clear and consistent mastery of writing. This SAT rubric was slightly adapted for source-based essays with the inclusion of evidence from the source as a criterion for writing quality. An example of the wording for an essay scored 6 using the independent essay rubric reads as follows: "A typical essay employs clearly appropriate examples, reasons, and other evidence to support its position." In the case of the dependent rubric, the wording was adjusted to state: "A typical essay utilizes clearly appropriate examples, reasons, and other evidence taken from the source text(s) to support its position." The objective was to align the two rubrics as closely as possible to ensure comparability between scores for independent and source-based essays.

As with the effectiveness ratings, expert raters employed a double-blind rating process, with 100 percent adjudication ensuring that each essay received holistic scores from two raters,

with a third rater intervening if necessary. Raters underwent training specific to the prompts and essay types used in the PERSUADE 2.0 corpus. Interrater agreement before adjudication indicated robust agreement among expert raters, with a weighted kappa of .75 and a Pearson r value of .75. The holistic scoring rubrics for both independent and source-based essays are provided in the Appendix. The overview of the variables included in the PERSUADE 2.0 corpus, along with detailed descriptions, is presented in Table 3-2.

Table 3-2 Labels and Descriptions of All Corpus Variables

VARIBLE	LABEL	DESCRIPTOR
Essay ID	essay_id_comp	Unique identifier assigned to anonymized essays in the corpus
Discourse Text	full_text	Complete text of each essay composed by participating students
Holistic Essay Scores	holistic_essay_score	Holistic scores (ranging from 1 to 6) assigned by human raters for each essay
Word Count	word_count	Total number of written words in each essay
Task Type	Task	Task type for each essay (categorized as independent and source-based task)
Gender	Gender	Demographic information indicating the gender of students (categorized as male or female)
Grade level	grade_level	Demographic information indicating the grade level of students (categorized as Grade 6, 8, 9, 10, 11, 12)
English Language Learning Status	ell_status	Demographic information indicating the non-native English language learner status of students (categorized as Yes or No)
Racial/Ethnicity Group	race_ethnicity	Demographic information indicating the race/ethnic group of students (categorized by White, Black/African American, Hispanic/Latino, Asian/Pacific Islander, American Indian/Alaskan Native, Multiracial/Other)
Socioeconomic Status	economically_disadvantaged	Demographic information indicating the socioeconomic status of students (categorized as economically disadvantaged or not economically disadvantaged)
Student disability status	student_disability_status	Demographic information indicating whether students have a disability requiring an IEP (categorized as having a disability or not having a disability)
Discourse Start Point	discourse_start	Starting point of the discourse within the essay for identifying each discourse type, denoted by a continuous numerical representation indicating the location of the first word occurrence within the entire essay
Discourse End Point	discourse_end	Ending point of the discourse within the essay for identifying each discourse type, denoted by a continuous numerical representation indicating the location of the last word occurrence within the entire essay
Discourse Type	discourse_type	Discourse components within the chosen excerpt of the full persuasive essay, classified into categories of lead, position, claim, counterclaim, rebuttal, evidence, concluding statement, or unannotated segments
Discourse Type Numbers	discourse_type_num	Numerical representation assigned to discourse elements within the chosen excerpt of the entire persuasive essay. For instance, if claims appear more than once in the full essay, they are designated as "claim 1," "claim 2," and so forth
Discourse Effectiveness	discourse_effectiveness	The effectiveness of discourse elements, as evaluated by human raters, categorized into three classifications: ineffective, adequate, and effective

3.6 DERIVED VARIABLES

The investigation of varied linguistic features across diverse student populations and the optimization of automated essay scoring systems to improve modeling performance serves as the impetus for this investigation. In this section, surface-level (i.e., microstructural) and higher meaning-level (i.e., macrostructural) attributes are introduced as derived variables. These attributes may play a pivotal role in discerning both shared characteristics and differences among student groups. Moreover, they may contribute significantly to the development and refinement of large language models for precise score prediction to evaluate GPT-revised essays. The models are expected to facilitate the implementation of AI revision systems that offer insightful revision suggestions to students on their original drafts.

3.6.1 Microstructural Elements

In this study, the primary NLP tool employed for textual analysis is the Coh-Metrix 3.0 language analysis tool, which was utilized to computationally evaluate each essay within the corpus. This tool facilitates the extraction of a diverse set of linguistic and textual features, as elucidated by McNamara and Graesser (2012). The software computes 108 linguistic and text-based features, which can be grouped into 11 broad categories reflecting various aspects of written language according to Graesser et al.'s (2004) manual. These categories encompass descriptive information, text easability principal component scores, referential cohesion, latent semantic analysis, lexical diversity, connectives, situation model, syntactic complexity, syntactic pattern density, word information, and readability.

It is important to note that the computation of these features is not arbitrary but rather grounded in educational assessment theories. These theories, rooted in discourse comprehension, provide substantive evidence that different dimensions of discourse, such as representations,

structures, strategies, and processes, can be systematically measured from surface to deeper levels of meaning (McNamara et al., 2010). For example, cohesion and coherence in written language can play a significant role in text comprehension. Texts with low cohesion are generally more easily understood than those with high cohesion; however, including cohesive elements or discourse markers in high-cohesion texts can enhance comprehension, particularly for less proficient readers (McCarthy et al., 2019). Numerous studies have demonstrated that Coh-Metrix indices, including referential cohesion, latent semantic analysis, and incidence of connectives, can be used to evaluate cohesion and coherence in essays (McNamara et al., 2014; McNamara & Graesser, 2012).

The selection of Coh-Metrix indices is also theoretically justified by the levels-of-language framework positing that a writer's proficiency is organized hierarchically across the subword/word, sentence, and discourse levels of language. Previous studies, as exemplified by the work of Wilson, Roscoe, and Ahmed (2017), have confirmed that nine Coh-Metrix measures extracted from middle school students' timed constructed responses align with word-, sentence-, and discourse-level skills, maintaining a consistent latent factor structure across grades.

In this study, Coh-Metrix (a research version rather than the openly available online tool) was used to evaluate the written texts from the constructed persuasive essays available in the PERSUADE 2.0 corpus. Each student essay underwent evaluation using the levels-of-language framework (refer to Troia et al., 2019 for details). This helps to establish proof of concept for the application of automated measures, grounded in the levels-of-language approach, to facilitate the assessment of persuasive essays authored by a large student population.

In this study, a range of measures were extracted from Coh-Metrix 3.0 to potentially evaluate writing skills across word, sentence, and discourse levels of language. The selection of

these measures was informed by their relevance as demonstrated in Chapter Two on writing measurements and analysis, as well as their plausible connection to the three language levels. Table 3-3 provides a summary of the selected indices and their interpretations.

Word-level indices included nine measures that captured the sophistication and diversity of students' word choice and vocabulary: mean syllables per word, measure of textual lexical diversity, mean word concreteness, mean word familiarity, mean word frequency, mean word hypernymy, mean word imageability, mean word meaningfulness, and mean word polysemy.

Sentence-level indices included 11 measures that evaluated the syntactic complexity and diversity of sentences within students' texts and syntactic cohesion: mean sentence length, standard deviation of sentence length, mean number of words before main verb, mean number of modifiers per noun phrase, mean minimal edit distance for all words, mean syntactic similarity for all sentences, incidence of agentless passive voice, noun phrase density, verb phrase density, negation density, and referential cohesion.

Discourse-level indices included 20 measures that examined features related to voice and perspective, semantic overlap, discourse cohesion, and coherence: incidence of first-person pronouns, incidence of second-person pronouns, incidence of third-person pronouns, argument overlap for all sentences, argument overlap for adjacent sentences, LSA overlap of adjacent paragraphs, LSA overlap of adjacent sentences, LSA given/new information, narrativity Z-score, deep cohesion, total words written, the incidence of all connectives, the incidence of causal connectives, the incidence of logical connectives, the incidence of adversative and contrastive connectives, the incidence of temporal connectives, the incidence of additive connectives, the incidence of positive connectives, the incidence of negative connectives, and readability.

3.6.3 Macrostructural Elements

3.6.3.1 Argument structure elements and effectiveness

The analysis of argument structure and its effectiveness is integral to the study. As aforementioned, the PERSUADE 2.0 corpus is characterized by its inclusion of annotations pertaining to argument structure and corresponding effectiveness scores (ineffective, adequate, effective) assigned by human raters to each identified element within a student persuasive essay. These features served as an important cluster of predictors for holistic scores of persuasive essay quality.

3.6.3.2 Content attributes

Constructing persuasive arguments represents a pivotal skill for students as it facilitates the articulation of their thoughts regarding pertinent issues, grounded in their topical knowledge derived either through source material or world knowledge (Duschl & Osborne, 2002). The entirety of their essays' content, particularly the arguments employed to articulate their stance, holds significance within the realm of literacy (Evagorou et al., 2023). This significance stems from the potential to reflect the students' depth of understanding and mastery of the subject matter, as well as their proficiency in synthesizing and analyzing information effectively. Given the persuasive nature of the essays, their content also serves to underscore language proficiency in conveying ideas with clarity and persuasiveness, both of which are imperative competencies in academic and professional domains for secondary-level students.

For this study, a text mining approach was utilized to compute the score point level using maximum cosine similarity across all score points and within each prompt to represent the content of the essay. This extraction method aligns with the approach utilized in Zupanc and Bosnic (2017) and Attali (2011). Cosine similarity measures the similarity between document

pairs by assessing the angle between their vectors, with values ranging from 0 to 1. A higher value signifies greater semantic and content similarity between the documents. Specifically, the cosine similarity of each essay was computed with human-assigned holistic scores ranging from 1 to 6, respectively, resulting in six content feature scores. The maximum cosine similarity score was retained as a metric representing the similarity between the target essay and the comparison essays. Because there are fifteen different prompts (with three of these prompts excluded because their corresponding texts did not contain all necessary features) in the corpus, cosine similarity scores were also calculated separately within each prompt. This approach yielded six content feature scores, which depict the content similarity between the target essay and the remaining essays across various score levels within the same prompt.

3.6.3.3 Style/Tone

Sentiment analysis is a field of measurement that investigates attitudes, opinions, and emotions directed towards a specific entity (Mostafa, 2017). It delves into the unique and evolving tones adopted by individual writers in relation to their subjects or topics (Shermis et al., 2013). For instance, when a writer intends to express disagreement with a scenario, sentiment analysis sheds light on the negativity, positivity, or neutrality inherent in the language used (Janda et al., 2019). In the context of automated essay scoring, sentiment analysis assumes a significant role, particularly in persuasive essays. Writers, especially in this genre, must articulate and substantiate their viewpoints on a subject, wherein the tone and manner of expressing textual sentiment significantly impact their writing.

When analyzing students' texts in the PERSUADE 2.0 corpus, well-crafted persuasive essays clearly articulate the opinion of the student writer and build support for their standpoint by expressing relevant ideas and concepts for their argument. Sentiment expressions unveil a

writer's judgments, evaluations, and feelings, likely used to express a preference for a particular position or to highlight the shortcomings of an alternative position. Furthermore, a considerable body of prior work has indicated variations in persuasiveness and tones across diverse student cohorts. Notably, students with disabilities encounter persistent difficulties in the domain of persuasive writing (Gleason, 1999; Nippold et al., 2005). For example, a persuasive essay may involve choosing a particular stance, developing structured paragraphs that substantiate facets of an opinion or argument, and persuading the reader to align with one's viewpoint (P.-L. Chuang & Yan, 2023; Englert et al., 2009). Specific difficulties exhibited by students with disabilities in crafting persuasive (argumentative or opinion) essays include adopting an inappropriate narrative writing style, using unsupported or nonexistent evidence, overlooking opposing perspectives, and presenting an argument that is consistent with the contrary standpoint (De La Paz et al., 2012; Gleason, 1999; Wissinger & De La Paz, 2020).

SEntimental ANalysis and Cognition Engine (hereafter, SEANCE) is a sentiment analysis tool used for academic discourse that relies on multiple preexisting sentiment, social-positioning, and cognition dictionaries. SEANCE offers a user-friendly interface, featuring 254 core indices and 20 component indices based on recent advancements in NLP sentiment analysis. Beyond the core indices, SEANCE allows for customization, enabling the inclusion of specific parts of speech and control for instances of negation. Notably, in SEANCE, any negated target word is disregarded within the relevant category. For instance, in processing the sentence "He is not happy," the term "happy" would not be counted as a positive emotion word. This methodology has demonstrated efficacy in excluding approximately 90% of negated words (Hutto & Gilbert, 2014).

SEANCE also incorporates the Stanford part of speech (POS) tagger (Toutanova et al., 2003) as implemented in Stanford CoreNLP (Manning et al., 2014). The POS tagger facilitates POS-tagged specific indices for nouns, verbs, and adjectives. This is crucial in sentiment analysis, as adjectives, verbs, and adverbs may convey unique aspects of sentiment more emphatically (Hatzivassiloglou & McKeown, 1997; Hu & Liu, 2004; Subrahmanian & Reforgiato, 2008). SEANCE provides reports on both POS and non-POS variables. Significantly, many vectors in SEANCE remain neutral regarding POS, allowing accurate processing of poorly formatted texts that may pose challenges for a POS tagger.

In this study, component scores extracted from the SEANCE analysis tool were used to measure aspects of emotion, cognition, and social order within the essays. These scores included the component score of negative adjectives (NRC negative adjectives, NRC disgust adjectives, NRC anger adjectives, GI negative adjectives, Lu Hui negative adjectives), the component score of social order (RC ethics verbs, GI need verbs, and RC rectitude words), the component score of positive adjectives (Lu Hui positive adjectives, Bader positive adjectives, GI positive adjectives, and Laswell positive affect adjectives), the component score of joy (NRC joy adjectives, NRC anticipation adjectives, NRC surprise adjective), and the component score of trust verbs (NRC trust verbs, NRC joy verbs, and NRC positive verbs). These scores collectively represent the tone of the essays.

Additionally, effective persuasive writing is accessible to readers because writers adhere to shared conventions for organization and signal their stance on a particular topic. In their research, Uccelli, Dobbs, and Scott (2013) discovered that the frequency of organizational markers and epistemic hedges significantly predicted high school students' persuasive essay writing quality. In academic persuasive writing, students are typically expected to employ

language markers that are more precise and suitable for the communicative context than those used in everyday interactions. However, these markers should not be overly specialized to the point of being discipline-specific (Bailey, 2007). To address this requirement, this study draws inspiration from the work of Islam, Xiao, and Mercer (2020), who compiled separate lists of hedge words, hedging phrases, and booster words. They also developed a rule-based algorithm that detects sentence-level hedges using these lexicons. Their NLP package and language marker lists are accessible on GitHub and align with the methodology employed in this study. Specifically, the stance markers were categorized into booster words (~128), discourse markers (~360), and hedge words (~80). Collectively, these stance markers were identified to enrich the analysis of students' persuasive writing.

The labels and descriptions of all the derived variables involved can be found in Table 3-3.

Table 3-3 Labels and Descriptions of the Study Variables

	Coh-Metrix Measure	Label	Brief Description
	Microstructural Elements		
	Word Level		
1	Mean syllables per word	DESWLsy	Word length, number of syllables, mean
2	MTLD	LDMTLD	Lexical diversity, MTLD, all words
3	Mean word concreteness	PCCNCz	Text Easability PC Word concreteness, z score
4	Mean word familiarity	WRDFAMc	Familiarity for content words, mean
5	Mean word frequency	WRDFRQc	Average word frequency for content words
6	Mean word hypernymy	WRDHYPnv	Hypernymy for nouns and verbs, mean
7	Mean word imageability	WRDIMGc	Imageability for content words, mean
8	Mean word meaningfulness	WRDMEAc	Meaningfulness, Colorado norms, content words, mean
9	Mean word polysemy	WRDPOLc	Polysemy for content words, mean
	Sentence Level		
10	Mean sentence length	DESSL	Sentence length, number of words, mean
11	Standard deviation of sentence length	DESSLd	Sentence length, number of words, standard deviation
12	Mean number of words before main verb	SYNLE	Left embeddedness, words before main verb, mean
13	Mean number of modifiers per noun phrase	SYNNP	Number of modifiers per noun phrase, mean
14	Mean minimal edit distance for all words	SYNMEDwrd	Minimal edit distance, all words
15	Mean syntactic similarity for all sentences	SYNSTRUTt	Sentence syntax similarity, all combinations, across paragraphs
16	Incidence of agentless passive voice	DRPVAL	Agentless passive voice density, incidence
17	Noun phrase density	DRNP	Noun phrase density, incidence
18	Verb phrase density	DRVP	Verb phrase density, incidence
19	Negation density	DRNEG	Negation density, incidence
20	Referential Cohesion	PCREFz	Text easability PC Referential cohesion, z score
	Discourse Level		
21	Incidence of first-person pronouns	WRDPRP1s	Use of first-person pronouns (personalization)
22	Incidence of second-person pronouns	WRDPRP2	Use of second-person pronouns (informality)
23	Incidence of third-person pronouns	WRDPRP3s	Use of third-person pronouns
24	Argument overlap for all sentences	CRFAOa	Argument overlap, all sentences, binary, mean
25	Argument overlap for adjacent sentences	CRFAO1	Argument overlap, adjacent sentences, binary, mean
26	LSA overlap of adjacent paragraphs	LSAPP1	Semantic similarity across paragraphs
27	LSA overlap of adjacent sentences	LSASS1	Semantic similarity across sentences
28	LSA given/new information	LSAGN	Semantic similarity of new text to prior text
29	Narrativity Z-score	PCNARz	Text Easability PC Narrativity, z score
30	Deep cohesion	PCDCz	Underlying conceptual cohesion
31	TWW	DESWC	Total Word Written
32	All connectives	CNCAll	All connectives incidence

Table 3-3 (cont'd)

33	Causal connectives	CNCCaus	Causal connectives incidence
34	Logical connectives	CNCLogic	Logical connectives incidence
35	Adversative and contrastive connectives	CNCADC	Adversative and contrastive connectives incidence
36	Temporal connectives	CNCTemp	Temporal connectives incidence
37	Additive connectives	CNCAdd	Additive connectives incidence
38	Positive connectives	CNCPos	Positive connectives incidence
39	Negative connectives	CNCNeg	Negative connectives incidence
40	Readability	RDFRE	Flesch Reading Ease

Macrostructural Elements

Structure

41	All elements	All_elements	Occurrences of all annotated elements
42	Lead	Lead	Occurrences of annotated lead
43	Position	Position	Occurrences of annotated position
44	Claim	Claim	Occurrences of annotated claim
45	Counterclaim	Counterclaim	Occurrences of annotated counterclaim
46	Rebuttal	Rebuttal	Occurrences of annotated rebuttal
47	Evidence	Evidence	Occurrences of annotated evidence
48	Concluding statement	Concluding Statement	Occurrences of annotated concluding statement
49	Total effectiveness score for all elements	All_effective_score	Effectiveness scores of all annotated elements
50	Effectiveness score for lead	Lead_effective	Effectiveness scores of annotated leads
51	Effectiveness score for position	Position_effective	Effectiveness scores of annotated positions
52	Effectiveness score for claim	Claim_effective	Effectiveness scores of annotated claims
53	Effectiveness score for counterclaim	Counterclaim_effective	Effectiveness scores of annotated counterclaims
54	Effectiveness score for rebuttal	Rebuttal_effective	Effectiveness scores of annotated rebuttals
55	Effectiveness score for evidence	Evidence_effective	Effectiveness scores of annotated evidences
56	Effectiveness score for concluding statement	Concluding_Statement_effective	Effectiveness scores of annotated concluding statement

Content

57	Content score 1	C1	Maximum similarity score between target essay and other essays scored as 1
58	Content score 2	C2	Maximum similarity score between target essay and other essays scored as 2
59	Content score 3	C3	Maximum similarity score between target essay and other essays scored as 3
60	Content score 4	C4	Maximum similarity score between target essay and other essays scored as 4
61	Content score 5	C5	Maximum similarity score between target essay and other essays scored as 5

Table 3-3 (cont'd)

62	Content score 6	C6	Maximum similarity score between target essay and other essays scored as 6
Tone/Style			
63	All stance markers	all_markers	All stance markers incidence
64	Booster words	booster_words	Booster words incidence
65	Discourse markers	discourse_words	Discourse markers incidence
66	Hedge words	hedge_words	Hedge words incidence
67	Negative adjectives component	Negative_adjectives_component	Component score of negative adjectives
68	Social order component	Social_order_component	Component score of social order
69	Positive adjectives component	Positive_adjectives_component	Component score of positive adjectives
70	Joy component	Joy_component	Component score of joy
71	Trust verbs component	Trust_verbs_component	Component score of trust verbs

3.7 RESEARCH DESIGN

3.7.1 Data Preparation

3.7.1.1 Subsample for Factor Analysis

To ensure data readiness and the accurate extraction of language features, a subset of 2,977 essays underwent screening and correction for common grammatical and mechanical errors by two trained human raters. Interrater reliability for this subset was evaluated using a two-way random-effects model with absolute agreement intraclass correlation (ICC). The resulting ICC values for all identified errors are reported in this subsection. No missing data was detected in the subset.

Normality assessments were performed using absolute values of skewness and kurtosis (z-skew and/or z-kurtosis $\geq$ |3.29| corresponding to an alpha level of 0.001, as suggested by Tabachnick & Fidell (2013), alongside P-P plots. Twenty-seven out of 71 variables exhibited a clear departure from normality: LDMTLD, PCCNCz, DESSL, DESSLd, SYNLE, SYNMEDwrd, DRPVAL, DRNEG, PCREFz, WRDPRP1s, WRDPRP2, WRDPRP3s, LSASS1, LSAGN, PCDCz, DESWC, CNCCaus, CNCADC, CNCTemp, CNCNeg, RDFRE, negative component score, joy component score, boosters, hedges, discourse markers, and all markers. It is worth noting that transforming these variables may result in the loss of their inherent scaling properties. Furthermore, these features are exclusively utilized for factor analysis, and the choice to fine-tune the estimator parameter (e.g., maximum likelihood with robust standard errors) to address non-normality issues has been made. Hence, it was decided not to transform these variables to achieve normality.

3.7.1.3 Subsample for Constructing and Validating Large Language Model

The remaining subset of the data ($n = \sim$22,000) was dedicated to subsequent phases of large language model construction and validation. Specifically, two columns within the subset of data were used for this purpose: the first column contains the full text data, which serves as the foundation for constructing the large language model. No missing data were identified within this subset. The second column used is holistic essay score, which serves as the ground truth for model validation by assessing modeling performance. This subset is important in addressing RQ3. The fine-tuned model resulting from this process is utilized to predict GPT-revised essays based on students' original essays. To revise the essays, students' original persuasive essays are input into GPT, prompting the system to revise the writing while considering research-based traits at the levels of language, tone/style, organization, and content. More specific and informative traits will be elaborated upon in Chapter Four based on findings from exploratory and confirmatory factor analyses. The prediction scores are then compared with the students' original scores to evaluate the increase in performance.

3.7.2 Analytic Procedure

The current study involves a secondary data analysis primarily employing social science quantitative research methodologies to investigate three key research questions. This study follows the principles of Design-Based Research (DBR; Barab & Squire, 2004). The ultimate goal is to examine the effectiveness of machine intelligence techniques within the realm of secondary writing assessment design. DBR is characterized by relating learning theories to practice and showing improvements of student learning in real-world settings (DBR Collective, 2003). While this study follows the DBR framework for its core processes, the objective of determining whether the collected and analyzed data lead to improved student learning through

the instructional tools will be a future endeavor. This aspect is further discussed in the future

direction section of Chapter Five.

Reeves & McKenney (2012) outlined three core processes of DBR: (a) analysis and

exploration, (b) design and construction, and (c) evaluation and reflection. This study follows the

three processes, and the implementation workflow is presented in Figure 3-4.
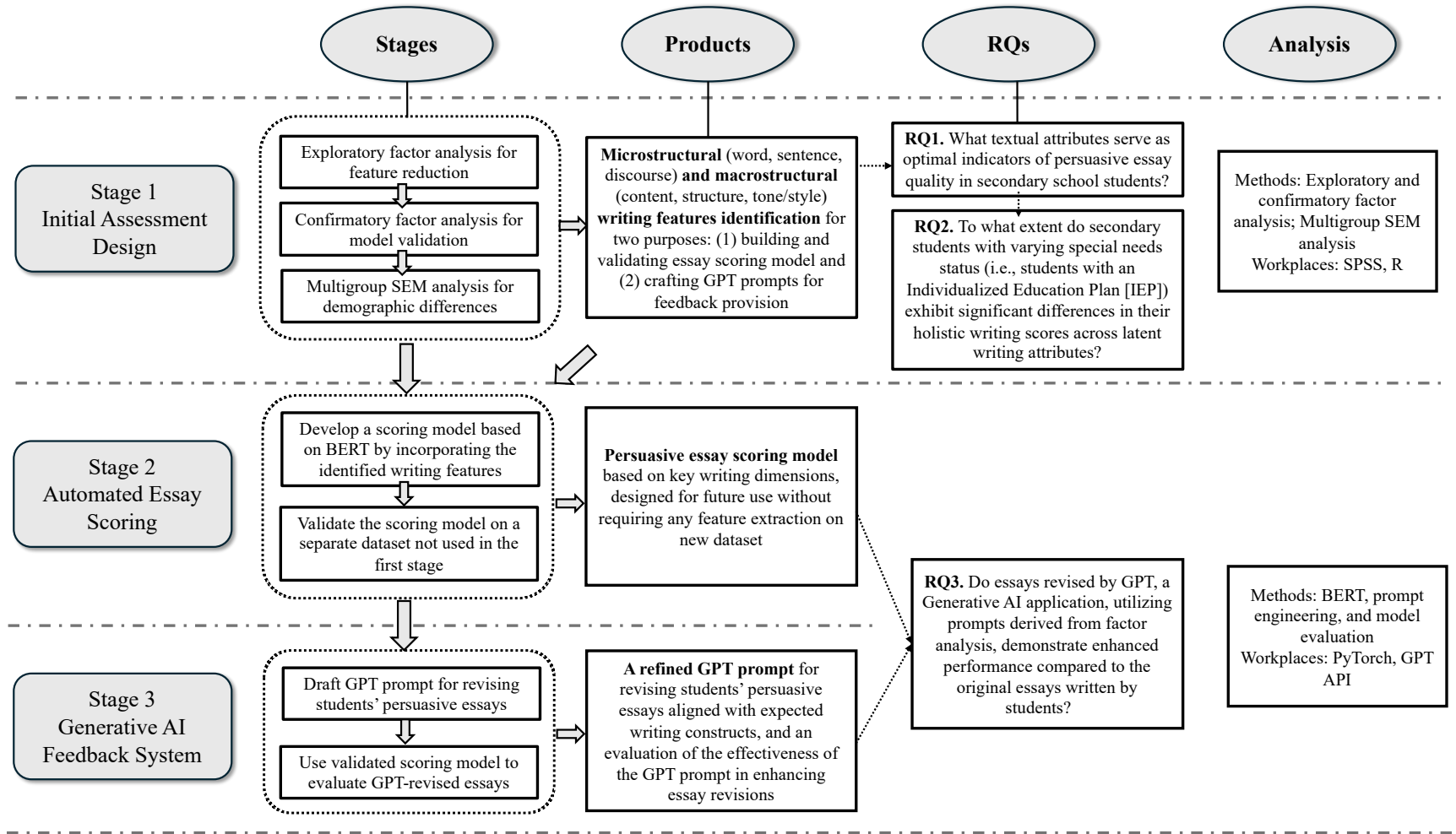
Figure 3-1 Implementation Workflow and Evaluation Methodology

3.7.2.1 RQ1: What textual attributes serve as optimal indicators of persuasive essay quality in secondary school students?

To address RQ1, I begin by employing exploratory factor analysis (EFA) to uncover underlying causal structures. This method allows me to discern the number and nature of latent variables that explain shared variability among observed indicators (C. D. Stapleton, 1997). In the context of the PERSUADE 2.0 corpus, six dimensions of writing are identified: language skills at the word, sentence, and discourse levels, as well as content, structure, and tone/style. These constructs encapsulate aspects of students' language use in response to specific prompts, thereby reflecting their writing ability through various writing-related measures in their essays. In this study, EFA with varimax rotation method is applied to (1) explain the common variance in a group of variables (e.g., DESWLsy, LDMTLD, PCCNCz, WRDFAMc, WRDFRQc, WRDHYPnv, WRDIMGc, WRDMEAc, WRDPOLc) by associating their observed scores with a reduced number of underlying latent factors (e.g., word level complexity) and (2) derive a loading matrix that demonstrates how each variable can be expressed as a linear combination of the common factors. These loadings (ranging from -1 to 1) are akin to correlations between the variables and the latent factor it represents. The study adheres to established guidelines (Costello & Osborne, 2019; Osborne, 2014) (Costello & Osborne, 2019; Osborne, 2014) recommending attributing items with absolute loadings of 0.5 or greater to the respective factor, with higher loadings indicating stronger associations. The varimax rotation method is employed to enhance the interpretability of the factors. Additionally, Kaiser-Meyer-Olkin (KMO) values for each category are reported to evaluate the adequacy of the data for EFA analysis.

Following the EFA, confirmatory factor analysis (CFA) is employed to establish evidence of construct validity within the hypothesized structural equation model. CFA facilitates the

concept proof of discriminant validity (Wilson et al., 2017), which assesses the differentiation among hypothesized levels of factors. In this study, CFA is applied across three distinct models, aligning with recommendations by Thompson (2000) and Rodgers et al. (2020). The first model (Model 1) is a baseline model where all writing variables are loaded onto a single factor (i.e., overall writing). The second model (Model 2) examines a two-factor model where all Coh-Metrix writing variables are loaded onto a microstructural factor, while other writing measures are loaded onto a macrostructural factor. The third model (Model 3) to be tested is a higher-order model that explores the possibility of six first-order factors (i.e., word, sentence, discourse, content, structure, tone/style) explained by the microstructural and macrostructural factors, respectively. Diagrams depicting each tested model are provided in Figure 3-1 for clarity and reference.
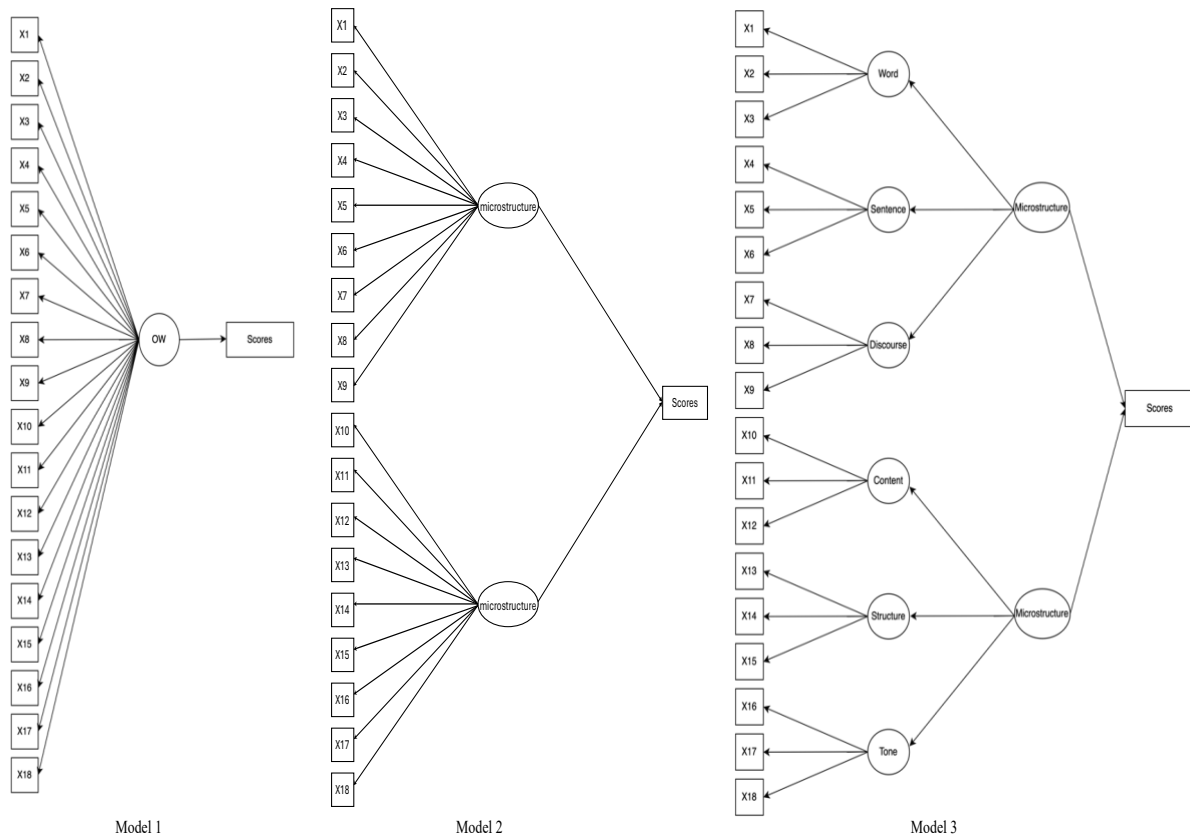
Figure 3-2 Diagrams for the One-factor, Two-factor, and Higher-order Models Evaluated

3.7.2.2 RQ2: To what extent do secondary students with different special needs status (i.e., students with and without an Individualized Education Plan [IEP]) exhibit significant differences in their holistic writing scores across latent writing attributes?

To address RQ2, multigroup structural equation modeling (multigroup SEM) (Byrne, 2016) facilitates the comparison of multiple samples across various population groups within the identified CFA model from RQ1. This analytical method assesses potential significant differences among various groups while assuming their equality and examining the concept of invariance. In the present study, the comparison focuses on group differences based on special education status.

3.7.2.3 RQ3: Do essays revised by GPT, a Generative AI application, utilizing prompts derived from factor analysis, demonstrate enhanced performance compared to original essays written by students?

To address RQ3, the study follows a two-step approach. The first step involves developing and fine-tuning a large language model customized specifically for this study. This model combines features derived from the MMFP with contextualized word embeddings from the Bidirectional Encoder Representations from Transformers (BERT) model. BERT has been widely acknowledged in prior research for its effectiveness in assessing students' responses to essay prompts and questions (Cochran et al., 2024; Z. Liu et al., 2023). It has emerged as the industry standard for a variety of NLP downstream tasks, e.g., score prediction. In addition, BERT excels as a contextualized embedding model for domain-specific tasks, which provides dynamic contextual representations for words within the essay to preserve both semantic and syntactic information (Wang et al., 2024). By fusing BERT-derived features with MMFP-derived features, the constructed language model aims to reflect students' comprehension of designated topics, thereby enhancing model performance and advancing our understanding of students' writing proficiency in persuasive writing tasks. This resulting language model is then utilized to predict students' persuasive writing scores.

In the second step, the GPT model is prompted to revise students' essays based on their original writing. A specific prompt, informed by insights from the MMFP model, is designed and integrated into the GPT to generate revised essays. It is hypothesized that the GPT-revised essays will demonstrate improvements across dimensions such as word usage, sentence structure, discourse organization, content coherence, structural clarity, and tone/style, all derived from the MMFP. These revised essays are then scored using the large language model developed in Step

1. Subsequently, the scores of the students' original essays are compared with the scores of the GPT-revised essays to evaluate any observed improvements.

3.7.3 Model Fit

To evaluate the fit of estimated EFA, CFA, and multigroup SEM models, two measures of absolute fit, the Chi-square test of model fit and the Standardized Root Mean Residual (SRMR), are employed. These measures gauge the extent to which the model's parameter estimates align with those derived from the observed data. The SRMR quantifies discrepancies between the model's implied correlations and the actual correlations observed in the raw data. It ranges from 0 to 1, with a value of 0 indicating a perfect fit and values below 0.08 suggesting good fit. Additionally, I evaluate model fit using two incremental/comparative measures: the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI). These metrics assess how well the postulated model fits relative to a more constrained baseline model. CFI and TLI values range from 0 to 1, with values approaching 0.9 indicating a relatively good model fit (Bentler, 1990). Acceptable fit values lie between 0.90 and 0.95, while values above 0.95 are considered indicative of excellent fit. Lastly, a parsimony-adjusted measure of fit, the Root Mean Square Error of Approximation (RMSEA) is utilized, which considers the number of estimated parameters in the model. RMSEA values below 0.05 indicate well-fitting models, while values exceeding 0.10 suggest poor fit.

To validate the constructed language model, performance was assessed using mean absolute error (MAE), the standard deviation of MAE, and R-squared (R^2), each averaged over a tenfold cross-validation. MAE measured the average magnitude of absolute differences between predicted and actual values, while the standard deviation of MAE indicated model stability. The low MAE values suggested that predicted values closely align with actual values, indicating a

95

high degree of accuracy. R^2, as a goodness-of-fit measure in regression models, represented the proportion of variance in the outcome variable—holistic essay scores—explained by the model.

3.7.4 Software

The analysis was conducted using the R Studio (R Core Team) lavvan (Rosseel, 2012) and "psych" package (Revelle, 2015) to conduct exploratory data analysis, EFA, CFA, and multigroup SEM with maximum likelihood estimation. The Python library Pytorch was utilized for training and evaluating BERT deep learning models. The pretrained BERT model can be accessed through the Huggingface library. GPT-3.5, accessed via the OpenAI API, is employed to revise students' essays.

Chapter 4: Findings and Discussion

In this chapter, I conduct a thorough analysis of the findings pertaining to the three research questions. Additionally, I interpret these results in relation to the existing literature. This section aims to offer an in-depth exploration of the results, emphasizing significant patterns and themes that have emerged from the data.

4.1 DESCRIPTIVE STATISTICS

Figure 4-1 illustrates the distribution of students across various grade levels sampled within each holistic score level (ranging from 1 to 6) of the PERSUADE 2.0 corpus. The observed data exhibit statistically significant differences from the expected values, as evidenced by Chi-square test for independence concerning grade level and holistic score, $\chi^2(20) = 6339.3$, $p < 0.001$. Hence, the null hypothesis is rejected, indicating that the data do not adhere to an equal distribution and that a relationship exists between grade level and holistic essay score. Grade level will not be considered as a grouping variable in the subsequent multigroup analysis and will be controlled as a covariate in further data analyses.
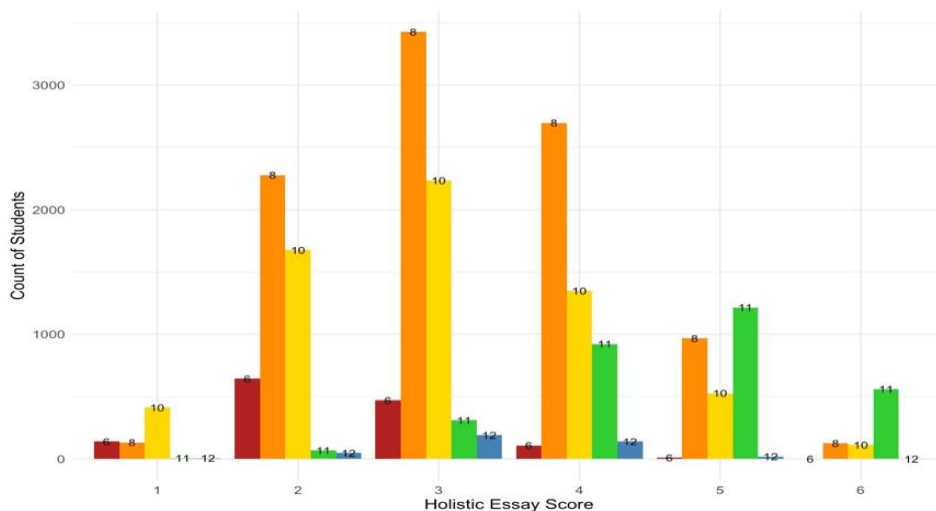
Figure 4-1 Distribution of Study Sample Across Holistic Score Levels of the PERSUADE 2.0 Corpus

Table 4-1 presents basic descriptive statistics pertaining to the writing measures under investigation in the present study.

Table 4-1 Means and (Standard Deviations) for Writing-Related Variables

	Grade 6	Grade 8	Grade 10	Grade 11	Grade 12
Holistic Score	1.92 (0.82)	3.06 (1.18)	2.59 (1.29)	4.85 (1.12)	2.76 (0.72)
DESWLsy	1.33 (0.08)	1.36 (0.10)	1.46 (0.11)	1.48 (0.12)	1.36 (0.07)
LDMTLD	68.0 (18.05)	66.9 (17.9)	75.2 (19.5)	73.2 (16.8)	54.7 (15.0)
PCCNCz	0.17 (1.05)	-0.35 (0.94)	-0.25 (0.855)	0.03 (0.62)	-0.03 (0.80)
WRDFAMc	582 (6.67)	583 (6.26)	581 (8.78)	584 (5.66)	590 (3.91)
WRDFRQc	2.58 (0.17)	2.60 (0.15)	2.46 (0.20)	2.45 (0.15)	2.71 (0.15)
WRDHYPnv	1.46 (0.17)	1.52 (0.18)	1.54 (0.18)	1.52 (0.12)	1.43 (0.16)
WRDIMGc	404 (23.67)	393 (21.3)	399 (18.7)	405 (13.6)	400 (19.8)
WRDMEAc	431 (13.2)	425 (15.9)	426 (15.4)	433 (11.3)	433 (13.7)
WRDPOLc	4.65 (0.44)	4.81 (0.57)	4.26 (0.51)	4.36 (0.47)	4.92 (0.51)
DESSL	20.0 (20.7)	22.1 (15.7)	24.1 (15.4)	24.4 (6.02)	29.1 (14.6)
DESSLd	8.9 (4.45)	11.5 (8.6)	11.4 (10.4)	10.2 (5.31)	16.6 (11.8)
SYNLE	3.54 (3.50)	4.28 (8.48)	4.40 (3.74)	5.14 (1.66)	4.85 (3.02)
SYNNP	0.67 (0.15)	0.65 (0.16)	0.75 (0.19)	0.74 (0.14)	0.57 (0.14)
SYNMEDwrd	0.85 (0.15)	0.82 (0.19)	0.86 (0.13)	0.85 (0.12)	0.85 (0.05)
SYNSTRUTt	0.10 (0.03)	0.07 (0.03)	0.08 (0.03)	0.08 (0.02)	0.07 (0.03)
DRPVAL	3.53 (4.61)	5.02 (5.73)	5.48 (5.08)	7.07 (5.02)	1.65 (2.48)
DRNP	358 (34.6)	342 (31.6)	337 (33.2)	332 (23.6)	352 (27.6)
DRVP	269 (34.9)	272 (41.3)	255 (40.6)	272 (32.7)	287 (40.2)
DRNEG	8.7 (8.6)	16.7 (10.3)	11.9 (9.1)	13.5 (6.9)	19.6 (10.1)
PCREFz	0.64 (1.18)	1.14 (1.30)	0.69 (1.19)	1.33 (1.09)	2.12 (1.43)
WRDPRP1s	21.65 (27.46)	17.91 (18.14)	8.45 (12.4)	7.37 (12.7)	14.4 (14.0)
WRDPRP2	34.95 (31.94)	22.7 (24.3)	11.3 (15.9)	6.52 (12.7)	23.9 (25.5)
WRDPRP3s	22.05 (29.11)	3.14 (8.03)	4.08 (8.93)	4.17 (10.1)	5.79 (13.5)
CRFAOa	0.56 (0.20)	0.60 (0.19)	0.61 (0.20)	0.72 (0.17)	0.74 (0.20)
CRFAO1	0.64 (0.20)	0.68 (0.19)	0.68 (0.20)	0.80 (0.15)	0.79 (0.18)
LSAPP1	0.29 (0.16)	0.34 (0.19)	0.40 (0.21)	0.54 (0.15)	0.46 (0.18)
LSASS1	0.17 (0.08)	0.24 (0.09)	0.26 (0.10)	0.29 (0.08)	0.30 (0.12)
LSAGN	0.30 (0.06)	0.34 (0.06)	0.34 (0.06)	0.37 (0.04)	0.39 (0.05)
PCNARz	1.12 (0.66)	1.10 (0.71)	0.56 (0.77)	0.60 (0.62)	1.45 (0.60)
PCDCz	0.87 (1.19)	1.72 (1.51)	1.10 (1.23)	1.67 (1.06)	2.67 (1.88)
DESWC	296 (102)	384 (199)	362 (174)	694 (216)	490 (216)
CNCAll	94 (21.6)	101 (23.1)	92.7 (20.5)	99.1 (15.5)	110 (24.2)
CNCCaus	28.6 (13.9)	36.4 (15.6)	30.7 (11.9)	36.7 (11.4)	45.1 (16.5)
CNCLogic	46.4 (17.8)	57.2 (18.3)	51.0 (17.4)	53.8 (12.9)	67.6 (18.4)
CNCADC	9.9 (8.5)	19.8 (11.0)	20.1 (9.87)	18.1 (7.01)	20.6 (10.0)
CNCTemp	13.6 (9.26)	14.9 (9.19)	13.7 (8.61)	16.0 (7.35)	11.9 (6.64)
CNCAdd	50.6 (16.1)	46.4 (15.5)	45.8 (14.7)	46.7 (12.1)	52.7 (13.1)
CNCPos	87.5 (22.3)	88.1 (22.6)	79.7 (18.9)	87.3 (15.3)	95.1 (23.0)
CNCNeg	8.6 (8.05)	15.3 (9.38)	15.0 (8.66)	13.8 (6.00)	18.3 (9.22)
RDFRE	75.1 (12.0)	69.9 (12.1)	59.7 (13.6)	56.8 (11.3)	62.7 (13.9)
Lead	0.30 (0.46)	0.53 (0.50)	0.51 (0.50)	0.92 (0.28)	0.54 (0.50)
Position	0.77 (0.42)	1.00 (0.13)	.93 (0.27)	1.00 (0.00)	1.00 (0.17)
Claim	2.34 (2.11)	2.80 (1.61)	2.53 (1.61)	4.31 (1.64)	3.87 (1.86)
Counterclaim	0.18 (0.50)	0.36 (0.63)	0.23 (0.57)	0.86 (0.86)	0.96 (1.34)
Rebuttal	0.13 (0.43)	0.27 (0.54)	0.15 (0.45)	0.69 (0.81)	0.25 (0.47)
Evidence	2.32 (1.42)	2.91 (1.24)	2.22 (1.25)	3.40 (1.15)	3.94 (1.61)
Concluding statement	0.56 (0.51)	0.83 (0.38)	0.78 (0.42)	0.98 (0.23)	0.82 (0.39)
Lead_effective	0.23 (0.47)	0.60 (0.73)	0.50 (0.67)	1.38 (0.71)	0.43 (0.50)
Position_effective	0.67 (0.54)	1.08 (0.58)	0.81 (0.62)	1.59 (0.48)	0.83 (0.31)
Claim_effective	0.70 (0.58)	1.38 (0.88)	0.85 (0.62)	1.59 (0.48)	0.83 (0.31)
Counterclaim_effective	0.13 (0.33)	0.30 (0.54)	0.20 (0.49)	0.96 (0.84)	0.45 (0.49)
Rebuttal_effective	0.09 (0.28)	0.23 (0.50)	0.14 (0.43)	0.78 (0.85)	0.22 (0.41)

Table 4-1 (cont'd)

|					
Evidence_effective	0.43 (0.44)	1.01 (0.61)	0.75 (0.60)	1.58 (0.53)	0.70 (0.31)
Concluding_statement_effective	0.50 (0.56)	0.83 (0.70)	0.77 (0.70)	1.54 (0.60)	0.63 (0.52)
C1	0.29 (0.06)	0.27 (0.01)	0.30 (0.08)	0.26 (0.04)	0.29 (0.04)
C2	0.33 (0.06)	0.33 (0.01)	0.32 (0.08)	0.37 (0.07)	0.36 (0.06)
C3	0.36 (0.06)	0.34 (0.02)	0.35 (0.09)	0.39 (0.07)	0.37 (0.06)
C4	0.37 (0.07)	0.36 (0.00)	0.37 (0.10)	0.42 (0.08)	0.36 (0.06)
C5	0.38 (0.08)	0.34 (0.10)	0.38 (0.11)	0.44 (0.07)	0.34 (0.06)
C6	0.39 (0.09)	0.34 (0.10)	0.37 (0.10)	0.45 (0.07)	0.34 (0.06)
All_markers	10.2 (7.6)	18.4 (12.2)	16.3 (10.3)	31.1 (12.3)	18.3 (9.33)
Booster_words	1.99 (2.11)	2.71 (3.00)	2.63 (2.40)	4.51 (3.52)	2.37 (2.30)
Discourse_markers	0.62 (1.11)	1.51 (2.06)	1.51 (1.83)	4.39 (3.15)	1.06 (1.65)
Hedge_words	7.6 (6.25)	14.2 (9.51)	12.2 (8.16)	22.2 (9.60)	14.9 (7.83)
Negative_adjectives_component	-0.58 (0.94)	-0.40 (0.89)	-0.07 (0.87)	-0.60 (0.61)	-0.67 (0.60)
Social_order_component	0.51 (0.15)	0.55 (0.18)	0.50 (0.15)	0.50 (0.13)	0.60 (0.21)
Positive_adjectives_component	0.13 (0.27)	0.24 (0.30)	0.10 (0.23)	0.21 (0.20)	0.20 (0.24)
Joy_component	1.24 (0.80)	1.08 (0.75)	0.93 (0.65)	0.70 (0.42)	0.91 (0.67)
Trust_verbs_component	0.15 (0.07)	0.16 (0.09)	0.19 (0.10)	0.25 (0.09)	0.18 (0.06)

Figure 4-2 displays a heatmap showing the unadjusted bivariate correlations between student demographic variables, essay scores, and writing measures. The X-axis represents various student demographic variables, including gender, grade level, English Language Learner (ELL) status, race/ethnicity, income background, disability status, and students' writing scores. The Y-axis includes writing measures at microstructural and macrostructural levels. Spearman or point-biserial correlation analyses were used for the categorical variables (e.g., gender, grade, ELL status, race/ethnicity, income background, and disability status) in relation to the writing elements. Pearson correlation analysis was employed for continuous variables, such as essay scores and writing measures. Only correlation coefficients with absolute values greater than 0.30 are presented in the Figure 4-2, as they indicate a moderate positive or negative linear relationship, consistent with standards in social science research (Bujang & Baharum, 2016; S. Crossley et al., 2023). P-values for these correlation coefficients are denoted with asterisks and interpreted as significantly correlated.

According to the Figure 4-2, grade level demonstrated significant correlations with several writing features, including mean syllables per word (DESWLsy), mean word frequency (WRDFRQc), mean word polysemy (WRDPOLc), incidence of first-person pronouns

(WRDPRP1s), incidence of second-person pronouns (WRDPRP2), LSA overlap of adjacent paragraphs (LSAPP1), and readability (RDFRE). Notably, these correlations are all positive. For instance, the positive and moderately strong correlation ($r = 0.39$, p < 0.001) between grade level and DESWLsy indicates that progression from lower grade levels (i.e., Grade 6) to higher grade levels (i.e., Grade 12) was moderately associated with an increase in mean syllables per word. These findings align with existing literature, suggesting that as students advance through secondary education and develop strategies for applying discourse-level knowledge, they tend to utilize more diverse lexical and syntactic choices (A. Truckenmiller et al., 2021), enhance their discourse cohesion and coherence (Sarmiento et al., 2024), and adopt genre-related rhetorical structures (W. Qin et al., 2024).

Regarding holistic essay scores, most of the structural and content-level features were found to be significantly and strongly correlated with paper quality. Some language features, such as mean syllables per word (DESWLsy), mean number of words before the main verb (SYNLE), LSA overlap of adjacent paragraphs (LSAPP1), LSA given/new information (LSAGN), and total word count (DESWC), along with some tone/style features (e.g., booster words, discourse markers, and hedge words), exhibit moderate to strong correlations with essay quality. These findings aligned with the PERSUADE corpus holistic rating rubric used to evaluate student essays, which emphasizes the importance of presenting clear claims, utilizing organized evidence to support positions, and demonstrating diverse and appropriate vocabulary choices and sentence structures. Existing literature suggests that the organizational features of argumentation and the quality of content are strong predictors of writing quality at the secondary levels, as these attributes enhance the clarity and persuasiveness of students' written work (S. A. Crossley et al., 2022; Taylor et al., 2019; Uccelli et al., 2013).
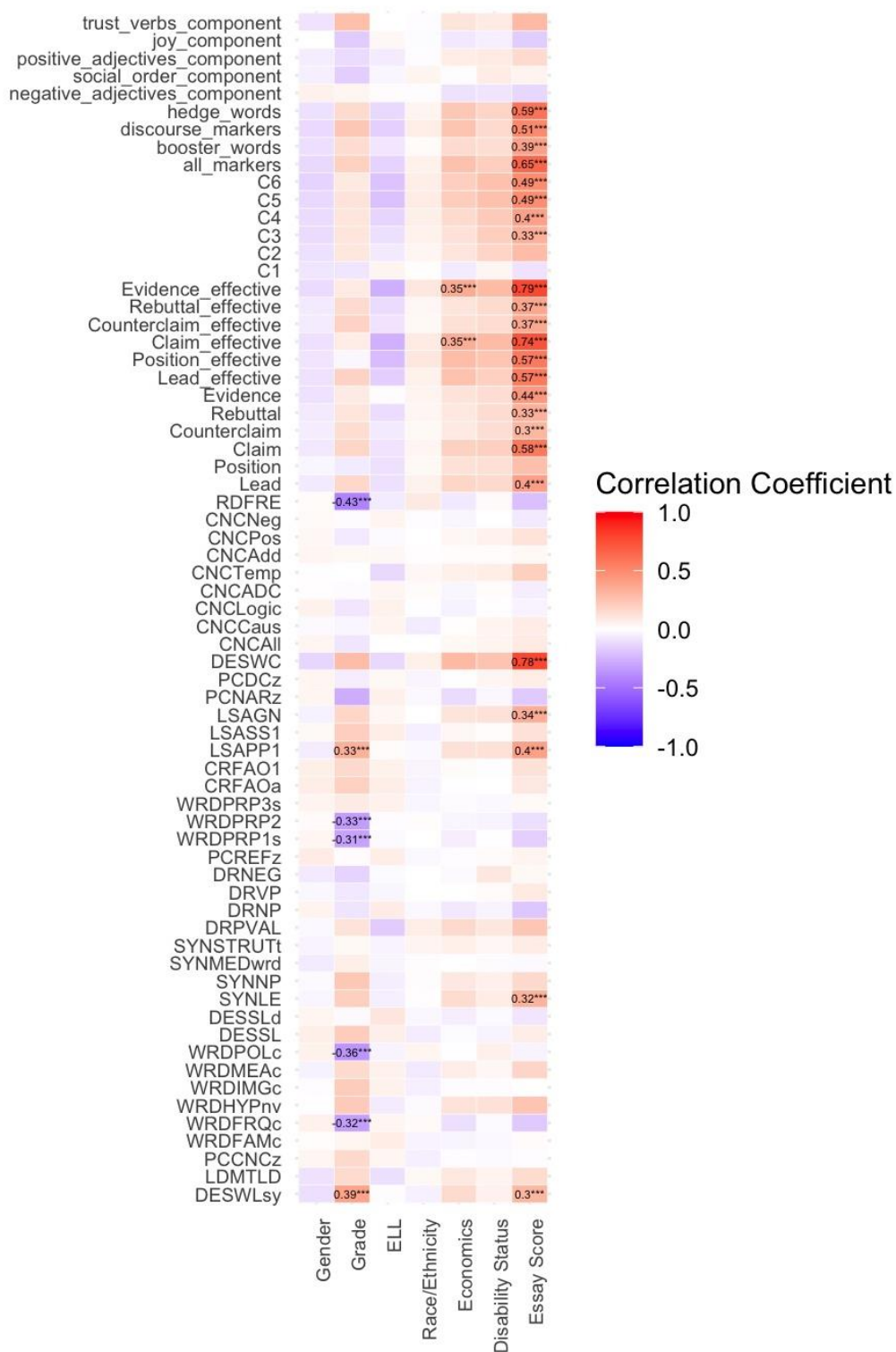
Figure 4-2 Correlation Heatmap of Writing-Related Variables, Demographic Information, and Essay Scores

Bivariate correlation coefficients among writing features at both the microstructural and microstructural levels were examined (see Table 4-2). As mentioned in Chapter 3 under the subsection of Research Design, normality assessments were conducted for all study variables, revealing that 27 writing-related variables exhibited non-normal distributions. To account for the non-normality, a weighted least square (WLS) estimator was employed for the analyses to extract factors and compute robust standard errors (Kline, 2015). This approach is recommended when the assumption of multivariate normality is violated according to Fabrigar et al. (1999), as was the case in this dataset, where many variables demonstrated skewed distributions.

Table 4-2 Correlations Among All Study Writing Related Variables in the Corpus

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1. DESWLsy	1.00	0.45	0.03	-0.46	-0.73	0.47	0.22	0.12	-0.51	-0.10	-0.19	0.03	0.57	0.08	0.20	0.23	0.03	-0.37	-0.29	-0.26	-0.38	-0.32	-0.04
2. LDMTLD	0.45	1.00	0.09	-0.38	-0.53	0.20	0.16	0.03	-0.37	-0.07	-0.08	0.02	0.38	0.19	-0.02	0.13	0.02	-0.34	-0.16	-0.65	-0.12	-0.16	0.00
3. PCCNCz	0.03	0.09	1.00	0.03	-0.30	0.43	0.72	0.61	-0.04	0.39	0.14	0.26	0.20	-0.28	-0.20	0.03	0.06	-0.12	-0.34	0.16	-0.14	-0.06	0.05
4. WRDFAMc	-0.46	-0.38	0.03	1.00	0.66	-0.23	-0.17	0.30	0.39	0.05	0.09	-0.01	-0.35	-0.08	-0.05	-0.26	-0.30	0.50	0.15	0.30	0.22	0.25	0.04
5. WRDFRQc	-0.73	-0.53	-0.30	0.66	1.00	-0.54	-0.48	-0.15	0.61	0.06	0.13	-0.04	-0.60	-0.12	-0.13	-0.28	-0.12	0.51	0.34	0.37	0.35	0.29	0.01
6. WRDHYPnv	0.47	0.20	0.43	-0.23	-0.54	1.00	0.60	0.37	-0.16	-0.05	-0.07	0.01	0.46	-0.01	0.09	0.17	0.09	-0.21	-0.18	-0.05	-0.26	-0.24	-0.04
7. WRDIMGc	0.22	0.16	0.72	-0.17	-0.48	0.60	1.00	0.55	-0.29	-0.05	-0.07	0.00	0.34	0.04	0.10	0.17	0.16	-0.29	-0.22	-0.12	-0.21	-0.23	-0.01
8. WRDMEAc	0.12	0.03	0.61	0.30	-0.15	0.37	0.55	1.00	0.00	-0.01	-0.04	0.00	0.08	-0.05	0.04	-0.06	-0.07	0.15	-0.24	0.03	-0.02	0.09	0.05
9. WRDPOLc	-0.51	-0.37	-0.04	0.39	0.61	-0.16	-0.29	0.00	1.00	0.05	0.11	-0.05	-0.43	-0.12	-0.14	-0.15	-0.10	0.48	0.16	0.30	0.21	0.26	-0.01
10. DESSL	-0.10	-0.07	0.39	0.05	0.06	-0.05	-0.05	-0.01	0.05	1.00	0.55	0.62	-0.04	-0.27	-0.50	-0.08	-0.01	0.01	-0.03	0.33	0.03	-0.01	0.07
11. DESSLd	-0.19	-0.08	0.14	0.09	0.13	-0.07	-0.07	-0.04	0.11	0.55	1.00	0.08	-0.10	-0.09	-0.44	-0.10	-0.04	0.04	0.07	0.23	0.10	0.04	0.06
12. SYNLE	0.03	0.02	0.26	-0.01	-0.04	0.01	0.00	0.00	-0.05	0.62	0.08	1.00	0.05	-0.13	-0.15	0.02	0.02	-0.04	-0.04	0.10	-0.01	-0.04	0.04
13. SYNNP	0.57	0.38	0.20	-0.35	-0.60	0.46	0.34	0.08	-0.43	-0.04	-0.10	0.05	1.00	0.08	0.10	0.19	-0.08	-0.57	-0.27	-0.32	-0.33	-0.27	-0.01
14. SYNMEDwrd	0.08	0.19	-0.28	-0.08	-0.12	-0.01	0.04	-0.05	-0.12	-0.27	-0.09	-0.13	0.08	1.00	0.10	0.04	0.03	-0.11	-0.04	-0.54	-0.05	-0.06	0.02
15. SYNSTRUTt	0.20	-0.02	-0.20	-0.05	-0.13	0.09	0.04	0.04	-0.14	-0.50	-0.44	-0.15	0.10	0.10	1.00	0.08	0.10	-0.06	-0.06	-0.21	-0.10	-0.10	-0.02
16. DRPVAL	0.23	0.13	0.03	-0.26	-0.28	0.17	0.17	-0.06	-0.15	-0.08	-0.10	0.02	0.19	0.04	0.08	1.00	-0.02	-0.07	-0.04	-0.12	-0.18	-0.24	-0.09
17. DRNP	0.03	0.02	0.06	-0.30	-0.12	0.09	0.16	-0.10	-0.10	-0.04	-0.04	0.02	-0.08	0.03	0.10	-0.02	1.00	-0.49	-0.19	-0.04	0.01	-0.08	0.02
18. DRVP	-0.37	-0.34	-0.12	0.50	0.51	-0.21	-0.29	0.15	0.48	0.01	0.04	-0.04	-0.57	-0.11	-0.06	-0.07	-0.49	1.00	0.23	0.29	0.20	0.22	-0.03
19. DRNEG	-0.29	-0.16	-0.34	0.15	0.34	-0.18	-0.22	-0.24	0.16	-0.03	0.07	-0.04	-0.27	-0.04	-0.06	-0.04	-0.19	0.23	1.00	0.11	0.16	0.05	-0.05
20. PCREFz	-0.26	-0.65	0.16	0.30	0.37	-0.05	-0.12	0.03	0.30	0.33	0.23	0.10	-0.32	-0.54	-0.21	-0.12	-0.04	0.29	0.11	1.00	0.07	0.12	0.01
21. WRDPRP1s	-0.38	-0.12	-0.14	0.22	0.35	-0.26	-0.21	-0.02	0.21	0.03	0.10	-0.01	-0.33	-0.05	-0.10	-0.18	0.01	0.20	0.16	0.07	1.00	0.15	0.08
22. WRDPRP2	-0.32	-0.16	-0.06	0.25	0.29	-0.24	-0.23	0.09	0.26	-0.01	0.04	-0.04	-0.27	-0.06	-0.10	-0.24	-0.08	0.22	0.05	0.12	0.15	1.00	-0.01
23. WRDPRP3s	-0.04	0.00	0.05	0.04	0.01	-0.04	-0.01	0.05	-0.01	0.07	0.06	0.04	-0.01	0.02	-0.02	-0.09	0.02	-0.03	-0.05	0.01	0.08	-0.01	1.00
24. CRFAOa	-0.05	-0.36	0.17	0.15	0.13	0.02	-0.04	0.02	0.08	0.28	0.27	0.03	-0.17	-0.15	-0.29	-0.05	-0.04	0.13	-0.03	0.79	-0.05	0.02	-0.01
25. CRFAO1	-0.05	-0.37	0.16	0.16	0.12	0.03	-0.04	0.06	0.10	0.21	0.23	0.01	-0.15	-0.14	-0.25	-0.05	-0.04	0.14	-0.03	0.76	0.00	0.07	0.04
26. LSAPP1	0.29	-0.10	-0.06	-0.01	-0.16	0.17	0.05	0.03	-0.13	-0.10	-0.18	-0.01	0.10	0.08	0.17	0.11	-0.03	-0.01	-0.07	0.25	-0.25	-0.13	-0.04
27. LSASS1	0.09	-0.41	0.10	0.06	0.00	0.15	0.03	0.03	0.05	0.18	0.16	0.02	-0.05	-0.15	-0.12	-0.01	-0.02	0.07	-0.04	0.73	-0.14	-0.06	-0.01
28. LSAGN	0.09	-0.41	-0.16	0.13	0.08	0.10	0.05	0.01	0.08	-0.23	-0.17	-0.11	-0.07	0.00	0.20	0.01	-0.03	0.12	0.03	0.55	-0.11	0.00	-0.01
29. PCNARz	-0.73	-0.46	-0.22	0.48	0.71	-0.53	-0.41	-0.12	0.45	0.15	0.32	-0.08	-0.72	0.03	-0.31	-0.32	-0.16	0.54	0.40	0.45	0.46	0.35	0.10
30. PCDCz	-0.22	-0.19	0.03	0.32	0.31	-0.10	-0.22	0.04	0.35	0.22	0.26	0.04	-0.24	-0.11	-0.25	-0.09	-0.18	0.22	0.11	0.34	0.08	0.19	0.00
31. DESWC	0.25	0.09	0.02	0.05	-0.17	0.17	0.02	0.14	-0.08	0.02	-0.05	0.08	0.12	0.06	-0.02	0.11	-0.12	0.03	-0.05	0.07	-0.09	-0.04	0.10
32. CNCAll	-0.23	-0.15	0.18	0.28	0.25	-0.13	-0.17	0.03	0.27	0.19	0.20	0.05	-0.23	-0.08	-0.28	-0.09	-0.29	0.21	0.08	0.23	0.08	0.19	-0.01
33. CNCCaus	-0.08	-0.19	-0.05	0.23	0.22	-0.07	-0.19	0.10	0.32	0.09	0.07	0.01	-0.22	-0.10	-0.10	-0.05	-0.07	0.20	0.10	0.28	0.07	0.14	-0.02
34. CNCLogic	-0.32	-0.21	-0.02	0.34	0.37	-0.17	-0.23	-0.07	0.17	0.14	0.19	0.01	-0.27	-0.10	-0.21	-0.16	-0.28	0.22	0.24	0.26	0.14	0.19	-0.02
35. CNCADC	-0.10	0.09	0.06	0.03	0.06	-0.13	-0.12	-0.14	-0.04	0.08	0.13	0.00	-0.02	-0.01	-0.20	-0.02	-0.17	-0.03	0.16	-0.03	0.02	0.00	-0.03
36. CNCTemp	-0.03	0.11	0.09	0.03	-0.06	-0.02	-0.02	-0.08	-0.04	-0.01	-0.01	0.06	0.08	0.05	0.02	0.07	-0.12	-0.08	-0.07	-0.09	0.00	0.08	0.03
37. CNCAdd	-0.10	-0.05	0.25	0.10	0.09	-0.06	-0.06	0.02	0.05	0.18	0.21	0.02	-0.09	-0.03	-0.28	-0.12	-0.22	0.10	0.02	0.07	0.02	0.04	0.03
38. CNCPos	-0.21	-0.21	0.11	0.29	0.26	-0.12	-0.19	0.07	0.30	0.16	0.16	0.04	-0.26	-0.09	-0.22	-0.09	-0.26	0.27	0.05	0.27	0.09	0.22	-0.01
39. CNCNeg	-0.18	0.02	0.11	0.13	0.14	-0.13	-0.10	-0.06	0.01	0.10	0.15	0.00	-0.07	-0.02	-0.22	-0.08	-0.18	0.05	0.17	0.01	0.06	0.05	-0.01
40. RDFRE	-0.65	-0.27	-0.29	0.29	0.48	-0.30	-0.11	-0.08	0.33	-0.61	-0.41	-0.27	-0.39	0.11	0.32	-0.10	0.00	0.26	0.24	-0.13	0.26	0.24	-0.03
41. Lead	0.20	0.07	-0.05	0.01	-0.10	0.08	0.02	0.02	-0.08	-0.10	-0.12	0.00	0.10	0.08	0.07	0.14	0.03	-0.05	-0.05	-0.04	-0.01	-0.02	0.00
42. Position	-0.08	-0.07	-0.09	0.13	0.13	-0.04	-0.11	0.06	0.08	-0.04	-0.04	0.00	-0.14	-0.01	-0.01	0.02	-0.10	0.16	0.11	0.06	0.10	0.03	-0.04
43. Claim	0.12	-0.06	-0.02	0.13	-0.02	0.16	0.00	0.16	0.04	-0.10	-0.11	0.00	-0.01	0.00	0.10	0.05	-0.13	0.17	-0.01	0.12	-0.10	0.02	0.01
44. Counterclaim	0.02	0.03	0.02	0.06	0.02	0.08	0.04	0.04	0.00	0.00	0.00	0.00	-0.02	0.02	-0.03	0.06	-0.04	0.05	0.10	0.02	-0.03	-0.06	-0.03
45. Rebuttal	0.07	0.07	0.01	0.01	-0.03	0.08	0.04	0.03	-0.02	-0.02	-0.04	0.01	0.02	0.03	-0.03	0.11	-0.05	0.02	0.07	-0.02	-0.03	-0.09	-0.04
46. Evidence	0.06	0.04	-0.04	0.02	-0.02	0.08	0.03	0.07	0.01	-0.10	-0.09	-0.02	-0.01	0.04	0.11	0.05	-0.03	0.05	0.02	-0.02	-0.07	0.00	-0.03
47. Concluding_Statement	0.12	0.03	-0.13	0.01	-0.02	0.06	-0.06	0.04	-0.03	-0.17	-0.15	-0.05	0.01	0.04	0.11	0.07	-0.05	0.05	0.03	-0.01	-0.04	-0.02	-0.03
48. Lead_effective	0.28	0.13	-0.01	-0.03	-0.19	0.16	0.04	0.09	-0.11	-0.08	-0.15	0.03	0.15	0.08	0.06	0.17	-0.02	-0.04	-0.07	-0.06	-0.13	-0.06	0.00
49. Position_effective	0.18	0.07	-0.06	0.01	-0.07	0.15	-0.06	0.11	0.01	-0.08	-0.09	0.00	0.07	0.02	0.01	0.12	-0.13	0.09	0.03	-0.01	-0.02	0.00	-0.01
50. Claim_effective	0.24	0.07	-0.03	0.06	-0.11	0.24	-0.02	0.17	-0.01	-0.10	-0.12	0.01	0.08	0.03	0.00	0.15	-0.19	0.14	0.00	0.02	-0.05	-0.02	-0.02
51. Counterclaim_effective	0.13	0.08	0.05	0.03	-0.08	0.16	0.08	0.09	-0.06	-0.06	0.00	-0.01	0.01	0.04	0.03	-0.03	0.10	-0.08	0.04	0.04	0.00	-0.05	-0.11
52. Rebuttal_effective	0.17	0.11	0.03	0.00	-0.10	0.16	0.07	0.07	-0.07	-0.02	-0.05	0.02	0.08	0.05	-0.02	0.12	-0.06	0.01	0.02	-0.03	-0.06	-0.14	-0.03
53. Evidence_effective	0.28	0.10	-0.02	0.00	-0.17	0.24	0.01	0.17	-0.04	-0.10	-0.15	0.02	0.11	0.05	0.03	0.15	-0.17	0.10	-0.03	0.00	-0.08	-0.04	0.01
54. Concl_Sttm_effective	0.33	0.11	-0.03	-0.02	-0.20	0.22	0.02	0.12	-0.09	-0.11	-0.18	0.02	0.17	0.06	0.11	0.19	-0.13	0.02	-0.06	-0.01	-0.16	-0.10	-0.01
55. C1	-0.04	-0.06	0.00	-0.26	-0.07	0.04	0.09	-0.19	-0.08	-0.01	0.00	-0.01	-0.06	-0.01	-0.05	0.03	0.16	-0.12	0.00	0.06	-0.10	-0.04	-0.08
56. C2	0.04	-0.17	-0.04	-0.03	0.01	0.19	0.05	-0.06	0.10	-0.06	-0.06	-0.03	-0.10	-0.04	-0.04	0.10	-0.01	0.10	0.02	0.27	-0.11	-0.03	-0.16
57. C3	0.10	-0.10	-0.01	-0.11	-0.06	0.23	0.11	-0.07	0.06	-0.10	-0.10	-0.04	-0.03	-0.03	0.00	0.16	0.00	0.03	0.00	0.21	-0.16	-0.11	-0.18
58. C4	0.18	-0.07	-0.02	-0.16	-0.14	0.26	0.12	-0.06	0.00	-0.10	-0.13	-0.02	0.03	-0.01	0.02	0.21	-0.01	0.00	-0.04	0.18	-0.20	-0.16	-0.18
59. C5	0.23	-0.02	0.02	-0.17	-0.19	0.34	0.16	-0.03	-0.03	-0.11	-0.14	-0.01	0.11	-0.01	0.03	0.25	-0.03	-0.03	-0.04	0.14	-0.20	-0.22	-0.16
60. C6	0.22	0.02	0.03	-0.21	-0.21	0.33	0.18	-0.07	-0.06	-0.10	-0.13	-0.01	0.12	0.02	0.04	0.29	0.02	-0.07	-0.01	0.09	-0.21	-0.26	-0.14
61. all_markers	0.14	0.05	-0.07	0.18	-0.01	0.05	-0.10	0.17	-0.02	-0.01	-0.06	0.03	-0.03	0.02	0.00	0.05	-0.28	0.25	-0.01	0.10	-0.02	0.06	0.01
62. booster_words	0.13	0.12	-0.08	0.01	-0.07	-0.04	-0.10	0.04	-0.05	0.00	-0.03	0.02	0.06	0.08	-0.02	0.05	-0.12	0.00	-0.05	-0.03	0.00	0.05	0.05
63. discourse_markers	0.31	0.21	-0.04	-0.09	-0.23	0.11	-0.02	0.01	-0.17	0.00	-0.07	0.05	0.15	0.06	0.01	0.17	-0.11	-0.06	-0.05	-0.05	-0.10	-0.12	0.01
64. hedge_words	0.06	-0.03	-0.05	0.25	0.07	0.05	-0.10	0.20	0.04	-0.01	-0.06	0.02	-0.09	-0.01	0.01	0.00	-0.29	0.33	0.02	0.16	0.00	0.09	0.00
65. neg_adj_comp	-0.08	0.01	-0.01	-0.03	-0.01	-0.09	0.03	-0.10	-0.05	0.02	0.06	-0.02	-0.04	0.02	-0.03	0.00	-0.03	-0.03	0.16	-0.03	-0.03	-0.02	0.01
66. social_order_comp	-0.24	-0.17	-0.08	0.17	0.29	-0.06	-0.14	-0.04	0.33	0.02	0.03	-0.03	-0.22	-0.07	-0.05	-0.15	-0.03	0.27	0.11	0.14	0.07	0.12	-0.02
67. positive_adj_comp	-0.15	-0.14	-0.02	0.26	0.22	-0.01	-0.15	0.27	0.29	-0.01	0.03	-0.03	-0.19	-0.07	-0.09	-0.11	-0.27	0.36	0.03	0.10	0.17	0.20	-0.02
68. joy_comp	-0.28	-0.14	0.00	0.19	0.25	-0.15	-0.07	0.13	0.22	0.00	0.07	-0.05	-0.17	-0.04	-0.04	-0.10	-0.10	0.17	0.08	0.04	0.14	0.13	0.03
69. trust_verbs_comp	0.51	0.20	0.14	-0.25	-0.44	0.31	0.20	0.26	-0.32	-0.02	-0.11	0.04	0.28	0.01	0.11	0.16	-0.03	-0.12	-0.19	-0.10	-0.18	-0.16	0.02
70. all_elements	0.13	0.01	-0.03	0.11	-0.03	0.16	0.02	0.14	0.01	-0.11	-0.12	-0.01	0.00	0.04	0.10	0.10	-0.10	0.13	0.04	0.06	-0.10	-0.02	-0.02
71. all_effective_score	0.32	0.13	-0.01	0.01	-0.19	0.26	0.03	0.16	-0.08	-0.10	-0.15	0.03	0.14	0.06	0.04	0.20	-0.15	0.07	-0.02	-0.02	-0.11	-0.09	-0.02

Table 4-2 (cont'd)

	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
1. DESWLsy	-0.05	-0.05	0.29	0.09	0.09	-0.73	-0.22	0.25	-0.23	-0.08	-0.32	-0.10	-0.03	-0.10	-0.21	-0.18	-0.65	0.20	-0.08	0.12	0.02	0.07	0.06
2. LDMTLD	-0.36	-0.37	-0.10	-0.41	-0.41	-0.46	-0.19	0.09	-0.15	-0.19	-0.21	0.09	0.11	-0.05	-0.21	0.02	-0.27	0.07	-0.07	-0.06	0.03	0.07	0.04
3. PCCNCz	0.17	0.16	-0.06	0.10	-0.16	-0.22	0.03	0.02	0.18	-0.05	-0.02	0.06	0.09	0.25	0.11	0.11	-0.29	-0.05	-0.09	-0.02	0.02	0.01	-0.04
4. WRDFAMc	0.15	0.16	-0.01	0.06	0.13	0.48	0.32	0.05	0.28	0.23	0.34	0.03	0.03	0.10	0.29	0.13	0.29	0.01	0.13	0.13	0.06	0.01	0.02
5. WRDFRQc	0.13	0.12	-0.16	0.00	0.08	0.71	0.31	-0.17	0.25	0.22	0.37	0.06	-0.06	0.09	0.26	0.14	0.48	-0.10	0.13	-0.02	0.02	-0.03	-0.02
6. WRDHYPnv	0.02	0.03	0.17	0.15	0.10	-0.53	-0.10	0.17	-0.13	-0.07	-0.17	-0.13	-0.02	-0.06	-0.12	-0.13	-0.30	0.08	-0.04	0.16	0.08	0.08	0.08
7. WRDIMGc	-0.04	-0.04	0.05	0.03	-0.05	-0.41	-0.22	0.02	-0.17	-0.19	-0.23	-0.12	-0.02	-0.06	-0.19	-0.10	-0.11	0.02	-0.11	0.00	0.04	0.04	0.03
8. WRDMEAc	0.02	0.06	0.03	0.03	0.01	-0.12	0.04	0.14	0.03	0.10	-0.07	-0.14	-0.08	0.02	0.07	-0.06	-0.08	0.02	0.06	0.16	0.04	0.03	0.07
9. WRDPOLc	0.08	0.10	-0.13	0.05	0.08	0.45	0.35	-0.08	0.27	0.32	0.17	-0.04	-0.04	0.05	0.30	0.01	0.33	-0.08	0.08	0.04	0.00	-0.02	0.01
10. DESSL	0.28	0.21	-0.10	0.18	-0.23	0.15	0.22	0.02	0.19	0.09	0.14	0.08	-0.01	0.18	0.16	0.10	-0.61	-0.10	-0.04	-0.10	0.00	-0.02	-0.10
11. DESSLd	0.27	0.23	-0.18	0.16	-0.17	0.32	0.26	-0.05	0.20	0.07	0.19	0.13	-0.01	0.21	0.16	0.15	-0.41	-0.12	-0.04	-0.11	0.00	-0.04	-0.09
12. SYNLE	0.03	0.01	-0.01	0.02	-0.11	-0.08	0.04	0.08	0.05	0.01	0.01	0.00	0.06	0.02	0.04	0.00	-0.27	0.00	0.00	0.00	0.00	0.01	-0.02
13. SYNNP	-0.17	-0.15	0.10	-0.05	-0.07	-0.72	-0.24	0.12	-0.23	-0.22	-0.27	-0.02	0.08	-0.09	-0.26	-0.07	-0.39	0.10	-0.14	-0.01	-0.02	0.02	-0.01
14. SYNMEDwrd	-0.15	-0.14	0.08	-0.15	0.00	0.03	-0.11	0.06	-0.08	-0.10	-0.10	-0.01	0.05	-0.03	-0.09	-0.02	0.11	0.08	-0.01	0.00	0.02	0.03	0.04
15. SYNSTRUTt	-0.29	-0.25	0.17	-0.12	0.20	-0.31	-0.25	-0.02	-0.28	-0.10	-0.21	-0.20	0.02	-0.28	-0.22	-0.22	0.32	0.07	-0.01	0.10	-0.03	-0.03	0.11
16. DRPVAL	-0.05	-0.05	0.11	-0.01	0.01	-0.32	-0.09	0.11	-0.09	-0.05	-0.16	-0.02	0.07	-0.12	-0.09	-0.08	-0.10	0.14	0.02	0.05	0.06	0.11	0.05
17. DRNP	-0.04	-0.04	-0.03	-0.02	-0.03	-0.16	-0.18	-0.12	-0.29	-0.07	-0.28	-0.17	-0.12	-0.22	-0.26	-0.18	0.00	0.03	-0.10	-0.13	-0.04	-0.05	-0.03
18. DRVP	0.13	0.14	-0.01	0.07	0.12	0.54	0.22	0.03	0.21	0.22	0.22	-0.03	-0.08	0.10	0.27	0.05	0.26	-0.05	0.16	0.17	0.05	0.02	0.05
19. DRNEG	-0.03	-0.03	-0.07	-0.04	0.03	0.40	0.11	-0.05	0.08	0.10	0.24	0.16	-0.07	0.02	0.05	0.17	0.24	-0.05	0.11	-0.01	0.10	0.07	0.02
20. PCREFz	0.79	0.76	0.25	0.73	0.55	0.45	0.34	0.07	0.23	0.28	0.26	-0.03	-0.09	0.07	0.27	0.01	-0.13	-0.04	0.06	0.12	0.02	-0.02	-0.02
21. WRDPRP1s	-0.05	0.00	-0.25	-0.14	-0.11	0.46	0.08	-0.09	0.08	0.07	0.14	0.02	0.00	0.02	0.09	0.06	0.26	-0.10	0.10	-0.10	-0.03	-0.03	-0.07
22. WRDPRP2	0.02	0.07	-0.13	-0.06	0.00	0.35	0.19	-0.04	0.19	0.14	0.19	0.00	0.08	0.04	0.22	0.05	0.24	-0.02	0.03	0.02	-0.06	-0.09	0.00
23. WRDPRP3s	-0.01	0.04	-0.04	-0.01	-0.01	0.10	0.00	0.10	-0.01	-0.02	-0.02	-0.03	0.03	0.03	-0.01	-0.01	-0.03	0.00	-0.04	0.01	-0.03	-0.04	-0.03
24. CRFAOa	1.00	0.87	0.32	0.68	0.49	0.37	0.27	0.14	0.20	0.21	0.18	0.00	-0.04	0.10	0.21	0.01	-0.31	0.02	0.01	0.11	0.04	0.03	-0.03
25. CRFAO1	0.87	1.00	0.28	0.68	0.48	0.36	0.26	0.19	0.19	0.20	0.16	-0.01	-0.02	0.08	0.21	0.00	-0.26	0.04	0.02	0.13	0.05	0.03	0.01
26. LSAPP1	0.32	0.28	1.00	0.43	0.61	-0.16	0.01	0.41	-0.04	0.07	-0.07	-0.09	0.03	-0.09	0.00	-0.12	-0.16	0.26	0.09	0.30	0.10	0.10	0.17
27. LSASS1	0.68	0.68	0.43	1.00	0.71	0.14	0.23	0.11	0.12	0.20	0.11	-0.04	-0.06	0.02	0.16	-0.06	-0.31	0.01	0.00	0.11	-0.02	-0.04	-0.03
28. LSAGN	0.49	0.48	0.61	0.71	1.00	0.09	0.17	0.36	0.08	0.21	0.07	-0.10	0.01	-0.08	0.13	-0.11	0.05	0.18	0.09	0.33	0.07	0.05	0.20
29. PCNARz	0.37	0.36	-0.16	0.14	0.09	1.00	0.37	-0.12	0.32	0.23	0.42	0.09	-0.02	0.14	0.33	0.16	0.34	-0.16	0.13	-0.04	0.01	-0.03	-0.05
30. PCDCz	0.27	0.26	0.01	0.23	0.17	0.37	1.00	0.08	0.74	0.82	0.72	0.13	0.16	0.19	0.75	0.18	-0.07	-0.05	0.09	0.09	0.05	0.01	0.04
31. DESWC	0.14	0.19	0.41	0.11	0.36	-0.12	0.08	1.00	0.07	0.07	-0.08	0.16	0.02	0.10	0.39	-0.09	-0.23	0.40	0.12	0.49	0.26	0.28	0.43
32. CNCAll	0.20	0.19	-0.04	0.12	0.08	0.32	0.74	0.07	1.00	0.52	0.72	0.31	0.30	0.66	0.89	0.37	-0.02	-0.07	0.08	0.07	0.05	0.01	0.02
33. CNCCaus	0.21	0.20	0.07	0.20	0.21	0.23	0.82	0.07	0.52	1.00	0.45	-0.04	-0.10	-0.01	0.63	-0.02	-0.03	-0.01	0.08	0.08	0.04	0.01	0.03
34. CNCLogic	0.18	0.16	-0.07	0.11	0.07	0.42	0.72	-0.08	0.72	0.45	1.00	0.46	0.12	0.38	0.57	0.52	0.09	-0.12	0.11	0.00	0.06	0.01	-0.02
35. CNCADC	0.00	-0.01	-0.09	-0.04	-0.10	0.09	0.13	-0.08	0.31	-0.04	0.46	1.00	-0.01	0.42	-0.05	0.89	-0.02	-0.07	0.03	-0.12	0.10	0.08	-0.07
36. CNCTemp	-0.04	-0.02	0.03	-0.06	0.01	-0.02	0.16	0.02	0.30	-0.10	0.12	-0.01	1.00	0.03	0.30	-0.01	0.03	0.10	0.01	0.08	0.02	0.04	0.06
37. CNCAdd	0.10	0.08	-0.09	0.02	-0.08	0.14	0.19	0.10	0.66	-0.01	0.38	0.42	0.03	1.00	0.49	0.49	-0.10	-0.08	0.01	0.02	0.04	0.00	-0.02
38. CNCPos	0.21	0.21	0.00	0.16	0.13	0.33	0.75	0.39	0.63	0.89	0.57	-0.05	0.30	0.49	1.00	-0.02	0.00	-0.06	0.07	0.12	0.00	-0.03	0.04
39. CNCNeg	0.01	0.00	-0.12	-0.06	-0.11	0.16	0.18	-0.09	0.37	-0.02	0.52	0.89	-0.01	0.49	-0.02	1.00	0.03	-0.07	0.05	-0.12	0.11	0.07	-0.07
40. RDFRE	-0.31	-0.26	-0.16	-0.31	0.05	0.34	-0.07	-0.23	0.09	-0.02	0.03	-0.10	0.00	0.03	0.00	0.03	1.00	-0.07	0.10	-0.02	-0.03	-0.04	0.04
41. Lead	0.02	0.04	0.26	0.01	0.18	-0.16	-0.05	0.40	-0.07	-0.01	-0.12	-0.07	0.10	-0.08	-0.06	-0.07	-0.07	1.00	0.02	0.17	0.13	0.14	0.08
42. Position	0.01	0.02	0.09	0.00	0.09	0.13	0.09	0.12	0.08	0.08	0.11	0.03	0.01	0.01	0.07	0.05	0.10	0.02	1.00	0.24	0.09	0.07	0.19
43. Claim	0.11	0.13	0.30	0.11	0.33	-0.04	0.09	0.49	0.07	0.08	0.00	-0.12	0.08	0.02	0.12	-0.12	-0.02	0.17	0.24	1.00	0.08	0.09	0.49
44. Counterclaim	0.04	0.05	0.10	-0.02	0.07	0.01	0.05	0.26	0.05	0.04	0.06	0.10	0.02	0.04	0.00	0.11	-0.03	0.13	0.09	0.08	1.00	0.80	0.22
45. Rebuttal	0.03	0.03	0.10	-0.04	0.05	-0.03	0.01	0.28	0.01	0.01	0.01	0.08	0.04	0.00	-0.03	0.07	-0.04	0.14	0.07	0.09	0.80	1.00	0.16
46. Evidence	-0.03	0.01	0.17	-0.03	0.20	-0.05	0.04	0.43	0.02	0.03	-0.02	-0.07	0.06	-0.02	0.04	-0.07	0.04	0.08	0.19	0.49	0.22	0.16	1.00
47. Concluding_Statement	0.01	0.02	0.21	-0.01	0.21	-0.07	-0.01	0.30	-0.03	-0.04	-0.04	-0.04	0.02	-0.02	0.01	-0.05	0.04	0.20	0.20	0.25	0.13	0.15	0.17
48. Lead_effective	0.02	0.06	0.31	0.02	0.18	-0.21	-0.03	0.57	-0.03	0.01	-0.12	-0.07	0.14	-0.06	-0.01	-0.08	-0.14	0.75	0.11	0.27	0.15	0.19	0.15
49. Position_effective	0.02	0.05	0.18	-0.02	0.12	-0.04	0.07	0.38	0.05	0.06	-0.02	-0.03	0.07	-0.01	0.07	-0.04	-0.08	0.20	0.31	0.29	0.12	0.17	0.21
50. Claim_effective	0.07	0.11	0.29	0.05	0.23	-0.10	0.07	0.57	0.09	0.03	-0.01	-0.06	0.13	0.02	0.11	-0.07	-0.11	0.27	0.26	0.43	0.18	0.23	0.27
51. Counterclaim_effective	0.05	0.07	0.16	0.00	0.08	-0.06	0.02	0.36	0.03	0.02	0.01	0.06	0.04	0.03	0.00	0.06	-0.11	0.19	0.09	0.12	0.73	0.69	0.13
52. Rebuttal_effective	0.04	0.05	0.15	-0.01	0.06	-0.09	0.00	0.34	0.00	0.01	-0.02	0.04	0.05	0.00	-0.02	0.03	-0.12	0.18	0.07	0.12	0.61	0.77	0.13
53. Evidence_effective	0.07	0.12	0.30	0.04	0.22	-0.14	0.03	0.64	0.05	0.05	-0.06	-0.08	0.13	0.00	0.08	-0.09	-0.13	0.31	0.25	0.41	0.18	0.23	0.25
54. Concl_Sttm_effective	0.06	0.09	0.35	0.06	0.27	-0.23	-0.01	0.60	0.01	0.04	-0.11	-0.08	0.13	-0.04	0.04	-0.11	-0.16	0.35	0.19	0.39	0.15	0.22	0.25
55. C1	0.07	0.03	0.09	0.15	0.13	-0.03	-0.07	-0.04	-0.08	-0.04	-0.09	0.01	-0.07	-0.01	-0.06	-0.05	0.05	-0.09	-0.08	-0.08	-0.03	-0.03	0.05
56. C2	0.27	0.23	0.34	0.32	0.41	0.02	0.08	0.26	0.04	0.13	0.03	-0.03	-0.03	-0.01	0.09	-0.08	0.01	0.11	0.08	0.20	0.11	0.10	0.18
57. C3	0.23	0.18	0.35	0.28	0.39	-0.07	0.01	0.27	-0.01	0.06	-0.04	-0.03	-0.02	-0.03	0.03	-0.09	0.00	0.13	0.09	0.21	0.13	0.13	0.21
58. C4	0.22	0.18	0.37	0.28	0.39	-0.14	-0.02	0.31	-0.04	0.04	-0.09	-0.06	0.00	-0.05	0.01	-0.13	-0.05	0.16	0.09	0.24	0.12	0.14	0.21
59. C5	0.19	0.16	0.39	0.25	0.37	-0.20	-0.05	0.38	-0.05	0.00	-0.11	-0.05	0.03	-0.04	-0.01	-0.13	-0.09	0.18	0.11	0.28	0.14	0.17	0.22
60. C6	0.15	0.12	0.37	0.20	0.33	-0.21	-0.06	0.39	-0.02	-0.02	-0.12	-0.04	0.05	-0.08	-0.05	-0.12	-0.08	0.21	0.11	0.26	0.19	0.22	0.24
61. all_markers	0.15	0.18	0.33	0.10	0.31	0.04	0.12	0.76	0.13	0.11	0.06	-0.01	0.09	0.03	0.14	-0.01	-0.11	0.28	0.15	0.41	0.23	0.23	0.34
62. booster_words	0.04	0.08	0.22	0.00	0.16	-0.02	0.00	0.55	0.00	0.01	-0.06	0.02	0.08	-0.03	0.01	-0.02	-0.11	0.24	0.05	0.22	0.12	0.15	0.22
63. discourse_markers	0.07	0.08	0.27	0.05	0.16	-0.17	0.04	0.58	0.03	0.05	-0.03	0.10	0.13	-0.03	0.01	0.02	-0.24	0.27	0.09	0.25	0.25	0.30	0.19
64. hedge_words	0.16	0.19	0.30	0.12	0.32	0.10	0.15	0.68	0.16	0.13	0.10	-0.04	0.05	0.05	0.17	-0.01	-0.05	0.22	0.16	0.40	0.19	0.19	0.33
65. neg_adj_comp	-0.01	-0.04	-0.06	-0.04	-0.09	0.05	-0.04	-0.12	-0.01	0.02	0.06	-0.04	0.02	-0.06	0.08	0.03	-0.07	-0.06	-0.10	-0.01	-0.01	-0.01	-0.06
66. social_order_comp	0.04	0.04	-0.06	0.03	0.08	0.20	0.11	-0.01	0.09	0.11	0.11	0.01	-0.04	0.04	0.10	0.04	0.20	-0.05	0.03	0.03	0.02	-0.01	0.01
67. positive_adj_comp	0.00	0.02	-0.09	-0.01	0.00	0.23	0.21	0.02	0.20	0.19	0.19	-0.02	0.01	0.07	0.23	0.03	0.12	-0.08	0.15	0.08	0.04	0.04	0.03
68. joy_comp	-0.07	-0.05	-0.18	-0.05	-0.10	0.23	0.09	-0.21	0.09	0.05	0.14	0.03	-0.03	0.05	0.09	0.08	0.20	-0.16	0.03	-0.11	-0.03	-0.05	-0.06
69. trust_verbs_comp	-0.01	0.01	0.17	0.04	0.05	-0.36	-0.12	0.23	-0.10	-0.04	-0.18	-0.09	0.01	-0.03	-0.07	-0.13	-0.34	0.12	0.02	0.15	0.06	0.10	0.09
70. all_elements	0.06	0.10	0.32	0.04	0.32	-0.06	0.07	0.61	0.05	0.07	0.00	-0.07	0.10	0.00	0.08	-0.06	-0.02	0.33	0.30	0.81	0.51	0.48	0.75
71. all_effective_score	0.07	0.11	0.35	0.03	0.24	-0.19	0.02	0.69	0.04	0.05	-0.07	-0.05	0.14	-0.01	0.05	-0.06	-0.17	0.46	0.25	0.41	0.41	0.48	0.28

Table 4-2 (cont'd)

	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71
1. DESWLsy	0.12	0.28	0.18	0.24	0.13	0.17	0.28	0.33	-0.04	0.04	0.10	0.18	0.23	0.22	0.14	0.13	0.31	0.06	-0.08	-0.24	-0.15	-0.28	0.51	0.13	0.32
2. LDMTLD	0.03	0.13	0.07	0.07	0.08	0.11	0.10	0.11	-0.06	-0.17	-0.10	-0.07	-0.02	0.02	0.05	0.12	0.21	-0.03	0.01	-0.17	-0.14	-0.14	0.20	0.01	0.13
3. PCCNCz	-0.13	-0.01	-0.06	-0.03	0.05	0.03	-0.02	-0.03	0.00	-0.04	-0.01	-0.02	0.02	0.03	-0.07	-0.08	-0.04	-0.05	-0.01	-0.08	-0.02	0.00	0.14	-0.03	-0.01
4. WRDFAMc	0.01	-0.03	0.01	0.06	0.03	0.00	0.00	-0.02	-0.26	-0.03	-0.11	-0.16	-0.17	-0.21	0.18	0.01	-0.09	0.25	-0.03	0.17	0.26	0.19	-0.25	0.11	0.01
5. WRDFRQc	-0.02	-0.19	-0.07	-0.11	-0.08	-0.10	-0.17	-0.20	-0.07	0.01	-0.06	-0.14	-0.19	-0.21	-0.01	-0.07	-0.23	0.07	-0.01	0.29	0.22	0.25	-0.44	-0.03	-0.19
6. WRDHYPnv	0.06	0.16	0.15	0.24	0.16	0.16	0.24	0.22	0.04	0.19	0.23	0.26	0.34	0.33	0.05	-0.04	0.11	0.05	-0.09	-0.06	-0.01	-0.15	0.31	0.16	0.26
7. WRDIMGc	-0.06	0.04	-0.06	-0.02	0.08	0.07	0.01	0.02	0.09	0.05	0.11	0.12	0.16	0.18	-0.10	-0.10	-0.02	-0.10	0.03	-0.14	-0.15	-0.07	0.20	0.02	0.03
8. WRDMEAc	0.04	0.09	0.11	0.17	0.09	0.07	0.17	0.12	-0.19	-0.06	-0.07	-0.06	-0.03	-0.07	0.17	0.04	0.01	0.20	-0.10	-0.04	0.27	0.13	0.26	0.14	0.16
9. WRDPOLc	-0.03	-0.11	0.01	-0.01	-0.06	-0.07	-0.04	-0.09	0.08	0.10	0.06	0.00	-0.03	-0.06	-0.02	-0.05	-0.17	0.04	-0.05	0.33	0.29	0.22	-0.32	0.01	-0.08
10. DESSL	-0.17	-0.08	-0.08	-0.10	0.00	-0.02	-0.10	-0.11	-0.01	-0.06	-0.10	-0.10	-0.11	-0.10	-0.01	0.00	0.00	-0.01	0.02	-0.02	-0.01	0.00	-0.02	-0.11	-0.10
11. DESSLd	-0.15	-0.15	-0.09	-0.12	-0.01	-0.05	-0.15	-0.18	0.00	-0.06	-0.10	-0.13	-0.14	-0.13	-0.06	-0.03	-0.07	-0.06	0.06	0.03	0.03	0.07	-0.11	-0.12	-0.15
12. SYNLE	-0.05	0.03	0.00	0.01	0.01	0.02	0.02	0.02	-0.01	-0.03	-0.04	-0.02	-0.01	-0.01	0.03	0.02	0.05	0.02	-0.02	-0.03	-0.03	-0.05	0.04	-0.01	0.03
13. SYNNP	0.01	0.15	0.07	0.08	0.04	0.08	0.11	0.17	-0.06	-0.10	-0.03	0.03	0.11	0.12	-0.03	0.06	0.15	-0.09	-0.04	-0.22	-0.19	-0.17	0.28	0.00	0.14
14. SYNMEDwrd	0.04	0.08	0.02	0.03	0.03	0.05	0.05	0.06	-0.01	-0.04	-0.03	-0.01	-0.01	0.02	0.02	0.08	0.06	-0.01	0.02	-0.07	-0.07	-0.04	0.01	0.04	0.06
15. SYNSTRUTt	0.11	0.06	0.01	0.00	-0.03	-0.02	0.03	0.11	-0.05	-0.04	0.00	0.02	0.03	0.04	0.04	-0.02	0.01	0.01	-0.03	-0.05	-0.09	-0.04	0.11	0.10	0.04
16. DRPVAL	0.07	0.17	0.12	0.15	0.10	0.12	0.15	0.19	0.03	0.10	0.16	0.21	0.25	0.29	0.05	0.05	0.17	0.00	0.00	-0.15	-0.11	-0.10	0.16	0.10	0.20
17. DRNP	-0.05	-0.02	-0.13	-0.19	-0.08	-0.06	-0.17	-0.13	0.16	-0.01	0.00	-0.01	-0.03	0.02	-0.28	-0.12	-0.11	-0.29	-0.03	-0.03	-0.27	-0.10	-0.03	-0.10	-0.15
18. DRVP	0.05	-0.04	0.09	0.14	0.04	0.01	0.10	0.02	-0.12	0.10	0.03	0.00	-0.03	-0.07	0.25	0.00	-0.06	0.33	-0.03	0.27	0.36	0.17	-0.12	0.13	0.07
19. DRNEG	0.03	-0.07	0.03	0.00	0.04	0.02	-0.03	-0.06	0.00	0.02	0.00	-0.04	-0.04	-0.01	-0.01	-0.05	-0.05	0.02	0.16	0.11	0.03	0.08	-0.19	0.04	-0.02
20. PCREFz	-0.01	-0.06	-0.01	0.02	0.00	-0.03	0.00	-0.01	0.06	0.27	0.21	0.18	0.14	0.09	0.10	-0.03	-0.05	0.16	-0.03	0.14	0.10	0.04	-0.10	0.06	-0.02
21. WRDPRP1s	-0.04	-0.13	-0.02	-0.05	-0.06	-0.06	-0.08	-0.16	-0.10	-0.11	-0.16	-0.20	-0.20	-0.21	-0.02	0.00	-0.10	0.00	-0.03	0.07	0.17	0.14	-0.18	-0.01	-0.11
22. WRDPRP2	-0.02	-0.06	0.00	-0.02	-0.11	-0.14	-0.04	-0.10	-0.04	-0.03	-0.11	-0.16	-0.22	-0.26	0.06	0.05	-0.12	0.09	-0.02	0.12	0.20	0.13	-0.16	-0.02	-0.09
23. WRDPRP3s	-0.03	0.00	-0.01	-0.02	-0.03	-0.03	0.01	-0.01	-0.08	-0.03	-0.18	-0.18	-0.16	-0.14	0.01	0.05	0.01	0.00	0.01	-0.02	-0.02	0.03	0.02	-0.02	-0.02
24. CRFAOa	0.01	0.02	0.02	0.07	0.05	0.04	0.07	0.06	0.07	0.27	0.23	0.22	0.19	0.15	0.15	0.04	0.07	0.16	-0.01	0.04	0.00	-0.07	-0.01	0.06	0.07
25. CRFAO1	0.02	0.06	0.05	0.11	0.07	0.05	0.12	0.09	0.03	0.23	0.18	0.18	0.16	0.12	0.18	0.08	0.08	0.19	-0.04	0.04	0.02	-0.05	0.01	0.10	0.11
26. LSAPP1	0.21	0.31	0.18	0.29	0.16	0.15	0.30	0.35	0.09	0.34	0.35	0.37	0.39	0.37	0.33	0.22	0.27	0.30	-0.06	-0.06	-0.09	-0.18	0.17	0.32	0.35
27. LSASS1	-0.01	0.02	-0.02	0.05	0.00	-0.01	0.04	0.06	0.15	0.32	0.28	0.28	0.25	0.20	0.10	0.00	0.05	0.12	-0.04	0.03	-0.01	-0.05	0.04	0.04	0.03
28. LSAGN	0.21	0.18	0.12	0.23	0.08	0.06	0.22	0.27	0.13	0.41	0.39	0.39	0.37	0.33	0.31	0.16	0.16	0.32	-0.09	0.08	0.00	-0.10	0.05	0.32	0.24
29. PCNARz	-0.07	-0.21	-0.07	-0.10	-0.06	-0.09	-0.14	-0.23	-0.03	0.02	-0.07	-0.14	-0.20	-0.21	0.04	-0.02	-0.17	0.10	0.05	0.20	0.23	0.23	-0.36	-0.06	-0.19
30. PCDCz	-0.01	-0.03	0.04	0.07	0.02	0.00	0.03	-0.01	-0.07	0.08	0.01	-0.02	-0.05	-0.06	0.12	0.00	0.04	0.15	-0.04	0.11	0.21	0.09	-0.12	0.07	0.02
31. DESWC	0.30	0.57	0.38	0.57	0.36	0.34	0.64	0.60	-0.04	0.26	0.27	0.31	0.38	0.39	0.76	0.55	0.58	0.68	-0.12	-0.01	0.02	-0.21	0.23	0.61	0.69
32. CNCAll	-0.01	-0.03	0.05	0.09	0.03	0.00	0.05	0.01	-0.08	0.04	-0.01	-0.04	-0.05	-0.07	0.13	0.00	0.03	0.16	-0.01	0.09	0.20	0.09	-0.10	0.05	0.04
33. CNCCaus	0.03	0.01	0.06	0.08	0.02	0.01	0.05	0.04	-0.04	0.13	0.06	0.04	0.00	-0.02	0.11	0.01	0.05	0.13	-0.09	0.11	0.19	0.05	-0.04	0.07	0.05
34. CNCLogic	-0.04	-0.12	-0.02	-0.01	0.01	-0.02	-0.06	-0.11	0.04	0.03	-0.04	-0.09	-0.11	-0.12	0.06	-0.06	-0.03	0.10	0.02	0.11	0.19	0.14	-0.18	-0.04	-0.07
35. CNCADC	-0.04	-0.07	-0.03	-0.06	0.06	0.04	-0.08	-0.08	0.01	-0.03	-0.03	-0.06	-0.05	-0.04	-0.01	0.02	0.10	-0.04	0.06	0.01	-0.02	0.03	-0.09	-0.07	-0.05
36. CNCTemp	0.02	0.14	0.07	0.13	0.04	0.05	0.13	0.13	-0.07	-0.03	-0.02	0.00	0.03	0.05	0.09	0.08	0.13	0.05	-0.04	-0.04	0.01	-0.03	0.01	0.10	0.14
37. CNCAdd	-0.02	-0.06	-0.01	0.02	0.03	0.00	0.00	-0.04	-0.01	-0.01	-0.03	-0.05	-0.04	-0.08	0.03	-0.03	-0.03	0.05	0.02	0.04	0.07	0.05	-0.03	0.00	-0.01
38. CNCPos	0.01	-0.01	0.07	0.11	0.04	-0.02	0.08	0.04	-0.06	0.09	0.03	0.01	-0.01	-0.05	0.14	0.01	0.01	0.17	-0.06	0.10	0.23	0.09	-0.07	0.08	0.05
39. CNCNeg	-0.05	-0.08	-0.04	-0.07	0.06	0.03	-0.09	-0.11	-0.05	-0.08	-0.09	-0.13	-0.13	-0.12	-0.01	-0.02	0.02	-0.01	0.08	0.04	0.03	0.08	-0.13	-0.06	-0.06
40. RDFRE	0.04	-0.14	-0.08	-0.11	-0.11	-0.12	-0.13	-0.13	0.01	0.00	0.00	-0.05	-0.09	-0.08	-0.11	-0.11	-0.24	-0.05	0.03	0.20	0.12	0.20	-0.34	-0.02	-0.17
41. Lead	0.20	0.75	0.20	0.27	0.19	0.18	0.31	0.35	-0.09	0.11	0.13	0.16	0.18	0.21	0.28	0.24	0.27	0.22	-0.07	-0.05	-0.08	-0.16	0.12	0.33	0.46
42. Position	0.20	0.11	0.31	0.26	0.09	0.07	0.25	0.19	-0.08	0.08	0.09	0.09	0.11	0.11	0.15	0.05	0.09	0.16	-0.06	0.03	0.15	0.03	0.02	0.30	0.25
43. Claim	0.25	0.27	0.29	0.43	0.12	0.12	0.41	0.39	-0.08	0.20	0.21	0.24	0.28	0.26	0.41	0.22	0.25	0.40	-0.10	0.03	0.08	-0.11	0.15	0.81	0.41
44. Counterclaim	0.13	0.15	0.12	0.18	0.73	0.61	0.18	0.15	-0.03	0.11	0.13	0.12	0.14	0.19	0.23	0.12	0.25	0.19	-0.01	0.02	0.04	-0.03	0.06	0.51	0.41
45. Rebuttal	0.15	0.19	0.17	0.23	0.69	0.77	0.23	0.22	-0.03	0.10	0.13	0.14	0.17	0.22	0.23	0.15	0.30	0.19	-0.01	-0.01	0.04	-0.05	0.10	0.48	0.48
46. Evidence	0.17	0.15	0.21	0.27	0.16	0.13	0.25	0.25	0.05	0.18	0.21	0.21	0.22	0.24	0.34	0.22	0.19	0.33	-0.06	0.01	0.03	-0.06	0.09	0.75	0.28
47. Concluding_Statement	1.00	0.23	0.25	0.29	0.16	0.16	0.32	0.56	-0.03	0.17	0.20	0.22	0.25	0.23	0.27	0.17	0.20	0.25	-0.06	0.03	0.05	-0.08	0.11	0.30	0.40
48. Lead_effective	0.23	1.00	0.40	0.49	0.27	0.27	0.54	0.52	-0.10	0.14	0.16	0.21	0.28	0.30	0.42	0.32	0.42	0.35	-0.11	-0.05	-0.02	-0.20	0.21	0.39	0.71
49. Position_effective	0.25	0.40	1.00	0.59	0.22	0.23	0.56	0.47	-0.08	0.16	0.19	0.22	0.30	0.29	0.34	0.21	0.28	0.31	-0.12	0.01	0.16	-0.10	0.18	0.35	0.68
50. Claim_effective	0.29	0.49	0.59	1.00	0.34	0.33	0.77	0.60	-0.09	0.27	0.28	0.33	0.42	0.38	0.51	0.30	0.40	0.47	-0.14	0.04	0.18	-0.16	0.22	0.48	0.82
51. Counterclaim_effective	0.16	0.27	0.22	0.34	1.00	0.80	0.33	0.28	-0.05	0.17	0.19	0.20	0.24	0.29	0.30	0.18	0.35	0.24	-0.04	-0.01	0.06	-0.07	0.15	0.44	0.62
52. Rebuttal_effective	0.16	0.27	0.23	0.33	0.80	1.00	0.33	0.29	-0.05	0.15	0.17	0.19	0.24	0.29	0.28	0.19	0.34	0.22	-0.05	-0.03	0.03	-0.09	0.17	0.42	0.63
53. Evidence_effective	0.32	0.54	0.56	0.77	0.33	0.33	1.00	0.63	-0.09	0.25	0.27	0.33	0.43	0.40	0.54	0.33	0.45	0.49	-0.13	0.03	0.16	-0.15	0.26	0.46	0.83
54. Concl_Sttm_effective	0.56	0.52	0.47	0.60	0.28	0.29	0.63	1.00	-0.08	0.22	0.26	0.32	0.40	0.39	0.50	0.32	0.44	0.44	-0.13	-0.03	0.08	-0.16	0.26	0.45	0.77
55. C1	-0.03	-0.10	-0.08	-0.09	-0.05	-0.05	-0.09	-0.08	1.00	0.65	0.68	0.62	0.55	0.46	-0.10	-0.05	-0.06	-0.10	0.00	0.07	-0.04	-0.02	-0.10	-0.06	-0.11
56. C2	0.17	0.14	0.16	0.27	0.17	0.15	0.25	0.22	0.65	1.00	0.96	0.92	0.86	0.75	0.23	0.11	0.18	0.22	-0.14	0.12	0.09	-0.10	0.01	0.24	0.27
57. C3	0.20	0.16	0.19	0.28	0.19	0.17	0.27	0.26	0.68	0.96	1.00	0.97	0.92	0.83	0.22	0.12	0.19	0.21	-0.11	0.09	0.05	-0.12	0.05	0.27	0.30
58. C4	0.22	0.21	0.22	0.33	0.20	0.19	0.33	0.32	0.62	0.92	0.97	1.00	0.95	0.86	0.26	0.15	0.24	0.23	-0.13	0.05	0.04	-0.16	0.10	0.29	0.36
59. C5	0.25	0.28	0.30	0.42	0.24	0.24	0.43	0.40	0.55	0.86	0.92	0.95	1.00	0.90	0.29	0.17	0.30	0.25	-0.13	0.04	0.04	-0.18	0.13	0.34	0.46
60. C6	0.23	0.30	0.29	0.38	0.29	0.29	0.40	0.39	0.46	0.75	0.83	0.86	0.90	1.00	0.29	0.18	0.33	0.24	-0.11	-0.01	0.01	-0.14	0.16	0.35	0.46
61. all_markers	0.27	0.42	0.34	0.51	0.30	0.28	0.54	0.50	-0.10	0.23	0.22	0.26	0.29	0.29	1.00	0.64	0.62	0.95	-0.12	0.03	0.15	-0.10	0.16	0.50	0.58
62. booster_words	0.17	0.32	0.21	0.30	0.18	0.19	0.33	0.32	-0.05	0.11	0.12	0.15	0.17	0.18	0.64	1.00	0.43	0.43	-0.09	-0.03	0.07	-0.07	0.15	0.30	0.37
63. discourse_markers	0.20	0.42	0.28	0.40	0.35	0.34	0.45	0.44	-0.06	0.18	0.19	0.24	0.30	0.33	0.62	0.43	1.00	0.43	-0.12	-0.07	0.03	-0.15	0.26	0.36	0.54
64. hedge_words	0.25	0.35	0.31	0.47	0.24	0.22	0.49	0.44	-0.10	0.22	0.21	0.23	0.25	0.24	0.95	0.43	0.43	1.00	-0.10	0.07	0.17	-0.08	0.10	0.46	0.50
65. neg_adj_comp	-0.06	-0.11	-0.12	-0.14	-0.04	-0.05	-0.13	-0.13	0.00	-0.14	-0.11	-0.13	-0.13	-0.11	-0.12	-0.09	-0.12	-0.10	1.00	-0.07	-0.38	-0.07	-0.18	-0.09	-0.15
66. social_order_comp	0.03	-0.05	0.01	0.04	-0.01	-0.03	0.03	-0.03	0.07	0.12	0.09	0.05	0.04	-0.01	0.03	-0.03	-0.07	0.07	-0.07	1.00	0.13	0.01	-0.18	0.02	-0.01
67. positive_adj_comp	0.05	-0.02	0.16	0.18	0.06	0.03	0.16	0.08	-0.04	0.09	0.05	0.04	0.04	0.01	0.15	0.07	0.03	0.17	-0.38	0.13	1.00	0.52	0.11	0.07	0.12
68. joy_comp	-0.08	-0.20	-0.10	-0.16	-0.07	-0.09	-0.15	-0.16	-0.02	-0.10	-0.12	-0.16	-0.18	-0.14	-0.10	-0.07	-0.15	-0.08	-0.07	0.01	0.52	1.00	-0.05	-0.12	-0.19
69. trust_verbs_comp	0.11	0.21	0.18	0.22	0.15	0.17	0.26	0.26	-0.10	0.01	0.05	0.10	0.13	0.16	0.16	0.15	0.26	0.10	-0.18	-0.18	0.11	-0.05	1.00	0.17	0.29
70. all_elements	0.30	0.39	0.35	0.48	0.44	0.42	0.46	0.45	-0.06	0.24	0.27	0.29	0.34	0.35	0.50	0.30	0.36	0.46	-0.09	0.02	0.07	-0.12	0.17	1.00	0.59
71. all_effective_score	0.40	0.71	0.68	0.82	0.62	0.63	0.83	0.77	-0.11	0.27	0.30	0.36	0.46	0.46	0.58	0.37	0.54	0.50	-0.15	-0.01	0.12	-0.19	0.29	0.59	1.00

4.2 SAMPLING ADEQUACY

In order to perform subsequent factor analyses, the Kaiser-Meyer-Olkin (KMO) test for sufficiency of sample size for each category was evaluated. The KMO value was 0.68 and Bartlett's chi-square approximation was $\chi^2(2080) = 156181.4$, $p < .001$. A KMO value close to 1 indicates that the correlation pattern was compact enough to produce distinct and reliable factors. The resulting value of 0.68 was considered mediocre as it slightly under 0.70, a typical set minimum. It is recommended to make remedial action, which is to remove the variables with low values (typically below 0.50) and recalculate the KMO to see if it improves because the variables with low KMO values did not contribute to the common variance that other variables share, and retaining these variables in the factor analysis may introduce noise and inflate factor loadings (Howard, 2016). Thus, SYNLE (KMO = 0.36), SYNMEDwrd (KMO = 0.25), WRDPRP3s (KMO = 0.42), and CNCTemp (KMO = 0.27) were removed. The revised KMO value was found to be 0.79, which is considered as an average level of sampling adequacy for performing factor analysis on the dataset. By looking at the resultant Bartlett's sphericity test, $\chi^2(1830) = 143358.6$, $p < .001$, it was noted that newly acquired values were suitable to perform factor analysis.

4.3 ERROR CORRECTIONS

To ensure the generation of the most accurate writing features using the language analysis tool, it is essential to identify and correct spelling and grammatical errors in each writing sample. This correction process allows the samples to be accurately processed by Coh-Metrix, SEANCE, and semantic analysis tools, thereby enhancing the reliability of the extracted writing features. The total number of errors, encompassing both spelling and grammar, was recorded for each writing sample, and these combined errors were utilized to calculate inter-rater reliability between two raters.

The types of errors that require careful attention, as they may affect the ability of NLP tools to parse the text, are described as follows: (a) misspellings include obvious spelling mistakes (e.g., "becuase" should be corrected to "because"), omissions of parts of words (e.g., "in conclu" should be corrected to "in conclusion"), inappropriate acronyms (e.g., "u" should be corrected to "you"), and the incorrect separation of compound words (e.g., "class room" should be corrected to "classroom"); (b) mechanics errors include misplaced punctuation or the omission of necessary punctuation, which can affect the ability of Coh-Metrix to accurately count the number of sentences and related features within the text (e.g., "By asking multiple friends peers and even family members their thoughts for something that may or may not be important, it could help save someone from making a bad choice in the future." should be corrected to "By asking multiple friends, peers, and even family members their thoughts on something that may or may not be important, it could help save someone from making a bad choice in the future."); (c) messy codes are a frequent issue encountered by raters, as students typing on computers may inadvertently produce nonsensical characters that are difficult to interpret (e.g., "Ü Ü Ü Ü, @?@"). Raters systematically removed these extraneous codes.

It is noteworthy that errors in other aspects of sentence structure, such as capitalization errors, were not considered, as Coh-Metrix does not differentiate between lowercase and uppercase letters. Two independent raters conducted the corrections on the writing samples. I reviewed a total of 2,977 student essays, while a second rater—who is a doctoral student and former ESL teacher—marked corrections on 25% of randomly selected papers ($n = 745$) from the set I reviewed according to established guidelines. The percentage agreement regarding the identification of errors for correction was 91%. This figure was calculated by dividing the number of errors marked by the first rater by the average total number of errors identified by

both raters, as suggested by Polio (1997). A consensus was subsequently reached for the corrected versions of the papers. The finalized corrections were then input into Coh-Metrix, SEANCE, and content and semantic analysis tools to produce the writing-related variables.

4.4 MAIN ANALYSES

4.4.1 RQ1a. Persuasive Writing Features in Secondary School Students' Papers

Parallel analysis (PA) (Horn, 1965) and the scree plot (Cattell, 1966) were used to determine the appropriate number of factors to retain. PA compares the observed eigenvalues from the correlation matrix with those generated from uncorrelated normal variables. In this process, a factor is considered significant if its associated eigenvalue exceeds the 95th percentile of those obtained from random uncorrelated data (Glorfeld, 1995). The scree plot method provides a visual representation of where the eigenvalues sharply decline. Using both methods together helps avoid the risks of over- or under-extracting factors.

To address part of RQ1, all microstructural-level and macrostructural-level variables were loaded into PA software to determine the number of factors to retain. The scree plot from PA for the micro- and macrostructural level is presented in Figure 4-3. According to the PA results, sixteen factors should be retained, as these sixteen factors explained a substantial portion of the variance compared to random data. Collectively, the 16-factor solution accounted for the entirety of the variance in the Pearson correlation matrix, with the individual factors explaining 16%, 11%, 11%, 8%, 8%, 7%, 6%, 6%, 6%, 5%, 4%, 3%, 3%, 3%, 2%, and 2% of variance, respectively. However, this result may indicate an overfitting issue, where the algorithm identified statistically significant patterns that may not be practically meaningful. Based on the strong theoretical foundation outlined in Chapter 2 Literature Review – particularly models such as the levels of language framework, which supported a 3-factor model for microstructural

features, along with literature that categorized macrostructural features into content, structure, and tone/style – a 6-factor solution was imposed. This decision was made despite the screen plot and parallel analysis suggesting a higher number of factors, following the recommendations of Sürücü et al. (2024).
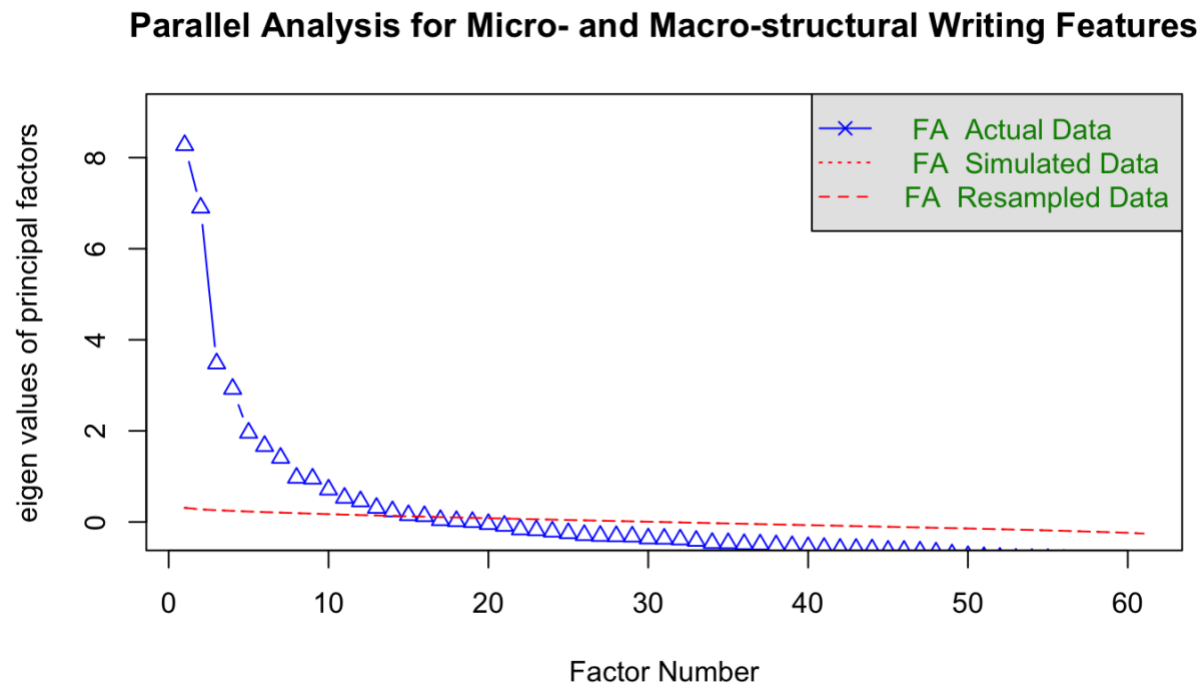
Figure 4-3 Scree Plot of Parallel Analysis for Microstructural Writing Features

After completing the preliminary factor analysis and determining the appropriate number of factors to retain, a Promax rotation was applied to the 6-factor solution, allowing the factors to correlate with each other. This solution accounted for the entire variance in the Pearson correlation matrix, with the six factors explaining 24%, 16%, 16%, 10%, 25%, and 10% of the total variance, respectively. Figure 4-4 presents the results of EFA, visually displaying the interpretation of the six-factor model. Only loadings with an absolute value greater than 0.50 were represented as edges, while those below the 0.50 threshold were excluded from further analyses.
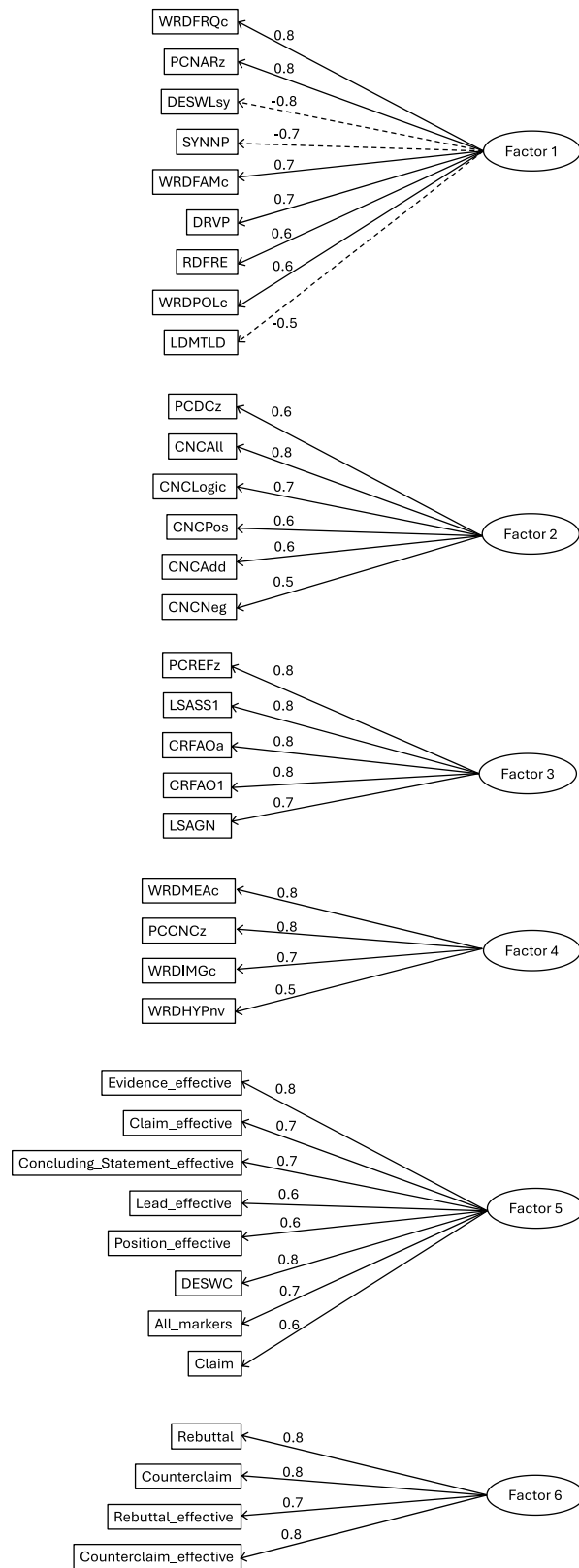
Figure 4-4 Exploratory Factor Analysis Plot

Table 4-3 presents the factor structure coefficients for the writing features examined in this study. The first factor comprised the following variables: DESWLsy, LDMTLD, WRDFAMc, WRDFRQc, WRDPOLc, SYNNP, DRPVAL, DRVP, WRDPRP1s, WRDPRP2, PCNARz, Readability, social_order_component, positive_adjectives_component, joy_component, and trust_verbs_component. Of these 16 variables, seven were excluded due to factor loadings below 0.5, indicating insufficient convergent validity as measured by average variance extracted. Retaining only variables with higher loadings was deemed beneficial for improving estimates of internal consistency reliability. The retained variables were DESWLsy, LDMTLD, WRDFAMc, WRDFRQc, WRDPOLc, SYNNP, DRVP, PCNARz, and Readability. Factor 1 was labeled **Lexical Proficiency**, as it captured both lexical complexity (e.g., mean syllables per word, mean word frequency, mean number of modifiers per noun phrase) and vocabulary diversity and depth (e.g., lexical diversity, mean word polysemy, word meaningfulness, mean word familiarity).

While some variables, such as Readability and Narrativity Z-score, might not seem directly aligned with Lexical Proficiency, they in fact underscore the role of lexical features in shaping word choice, text complexity, and discourse coherence, thus enhancing narrative effectiveness. Empirically, this outcome aligns with previous research (Aryadoust & Liu, 2015; Lu, 2017; Wilson et al., 2017) employing structural equation modeling to examine and validate NLP-derived writing features, thus informing nuanced analyses of text quality using hypothesized constructs. Our results showed that specific lexical features — especially those related to word information and vocabulary depth — had high loadings, indicating strong associations with the latent construct. Collectively, these lexical features were foundational to persuasive writing (i.e., explaining 24% of the variance). However, in contrast to findings from

other studies, readability and narrativity demonstrated substantial loadings of 0.64 and 0.76, respectively, under the Lexical Proficiency factor. This differs from previous research (Graesser et al., 2011; Nkhobo & Chaka, 2023; Plakans & Gebril, 2013), where these features are often associated with discourse-level elements. This discrepancy suggests that vocabulary richness and diversity notably impact readability and narrative depth in persuasive essays, supporting a more refined use of language to improve textual flow (McKeown et al., 2020).

The second factor included DESSL, DESSLd, SYNSTRUTt, DRNP, PCDCz, CNCAll, CNCCaus, CNCLogic, CNCADC, CNCAdd, CNCPos, and CNCNeg. After excluding variables with low factor loadings, the retained variables were PCDCz, CNCAll, CNCLogic, CNCAdd, CNCPos, and CNCNeg. This factor was labeled **Cohesive Devices**, capturing the integration of grammatical and lexical relationships both within and across sentences. This factor reflects cohesive device use, including the repetition of related words and concepts that enables readers to connect ideas at the sentence-to-sentence level (Graesser et al., 2011). Tortorelli (2020) similarly characterized cohesion as the effective linkage of ideas through referential and lexical ties. Our findings indicate that (1) cohesive devices such as reference, substitution, ellipsis, and conjunction are crucial for constructing persuasive discourse in writing and (2) the use of these cohesive devices collectively accounted for 16% of the variance in persuasive writing.

The Cohesive Devices factor thus represents explicit linguistic markers that serve to link individual sentences, maintain consistent references, reinforce thematic continuity, and facilitate a smooth flow of ideas (Kuo, 1995). The use of logical (e.g., thus, as a result, then, hence), additive (e.g., in fact, besides, in addition, furthermore), negative (e.g., neither/nor, but, however, though, conversely), and positive connectives (e.g., and, or, either/or, moreover, likewise, also) in persuasive writing provides linguistic leverage for signaling arguments and ultimately enhancing

writing quality (Taylor et al., 2019). Overall, this factor emphasizes the strategic use of discourse markers to connect ideas effectively.

The third factor included PCREFz, CRFAOa, CRFAO1, LSAPP1, LSASS1, LSAGN, C1, C2, and C6. After removing variables with low factor loadings, the retained variables were PCREFz, CRFAOa, CRFAO1, LSASS1, and LSAGN. The factor was labeled **Textual Coherence**, incorporating both micro- and macro-level cohesion. The key variables in this latent construct included LSA given/new information, LSA overlap of adjacent sentences/paragraphs, and deep cohesion, all of which contributed to coherence by establishing contextual ties rooted in shared knowledge between writer and reader invoking familiar or possible conceptual worlds (Widdowson, 1983).

Distinct from the Cohesive Devices factor, the Textual Coherence factor focuses on linking elements based on "thematic development, organization of information, or communicative purpose of the particular discourse" (Kuo, 1995, p.48). This factor emphasizes maintaining clarity across both small- and large-scale text segments. For example, LSA scores for adjacent/all sentences capture cohesion at a local or superficial level, such as repeating the same word choices across successive sentences (Roscoe et al., 2015). LSA given/new information evaluates the continuity of topic development and avoiding abrupt thematic shifts. This factor promotes reader comprehension of each sentence within the broader discourse. This finding is consistent with prior research in writing assessment, which posits that cohesion is not only syntactic but also semantic, each essential to effective essay composition (Ildikó Berzlánovich & Gisela Redeker, 2012; Witte & Faigley, 1981). An essay can be cohesively constructed yet lack coherence if it fails to represent shared frames of reference. Conversely, an essay may be coherent yet lack cohesive devices, limiting its ability to effectively communicate

connections between ideas. Hence, these findings underscore the complexity of cohesion and coherence, which together facilitate a deeper understanding of the interrelationships among ideas within persuasive writing.

The fourth factor consisted of PCCNCz, WRDHYPnv, WRDIMGc, and WRDMEAc, with DRNEG excluded due to a low loading value. This factor is challenging to categorize, yet I propose the label **Semantic Richness** to reflect its focus on qualities that enhance reader visualization and comprehension. Unlike Lexical Proficiency, which captures vocabulary complexity and diversity, Semantic Richness emphasizes the vivid, concrete, and meaningful aspects of word choice that facilitate sensory engagement and conceptual clarity. This factor underscores the ability of words to convey imagery and meaning, as well as their lexical hierarchy (such as hypernymy), which is essential in persuasive writing for grounding abstract concepts in relatable, vivid language.

The fifth factor included evidence_effective, claim_effective, concluding_statement_effective, lead_effective, position_effective, all_markers, claim, lead, position, evidence, concluding_statement, DESWC, booster_words, discourse_markers, and negative_adjectives_component. After removing the variables with low loadings, evidence_effective, claim_effective, concluding_statement_effective, lead_effective, position_effective, all_markers, claim, and DESWC were maintained. These retained variables distinctly represent the structural components of persuasive writing along with their respective effectiveness scores. Accordingly, this factor was labeled **Structural Effectiveness** to reflect its emphasis on the organization and quality of key argumentative elements within the text. This factor accounted for the largest proportion (i.e., 25%) of variance in persuasive writing,

indicating the variables involved under this category have a relatively large contribution to persuasive writing.

This finding is consistent with previous research (Kamimura, 2011; Tasya, 2022; Uccelli et al., 2013), which highlights that academic writing necessitates not only the mastery of language conventions but also advanced language forms and functions. De La Paz and colleagues (2012) noted that when composing argumentative essays, students often employed relevant evidence and argumentation strategies, effectively contextualizing and corroborating their evidence to enhance the overall quality of their writing. Likewise, Dobbs (2014) found that writers of all skill levels utilized organizational and stance markers to make claims and construct arguments. These findings indicate that effective persuasive writing is bolstered by the strategic use of structural and linguistic markers, as well as the effectiveness of discourse elements, all of which contribute to clarity, coherence, and engagement in the text.

The last factor, labeled **Refutation Quality**, included four variables: rebuttal, counterclaim, rebuttal_effective, and counterclaim_effective. These variables were distinguished from other elements of argumentation and accounted for 10% of the variance in explaining persuasive writing. This finding underscores the significance of argumentation skills in persuasive writing, particularly in how effectively writers address opposing viewpoints and bolster their arguments through rebuttals and counterclaims. Prior research (S. A. Crossley et al., 2014; F. Liu & Stapleton, 2014; J. Qin & Karabacak, 2010) has indicated that the incorporation of counterclaims and rebuttals in writing is often positively correlated with higher holistic scores for persuasive quality. Therefore, the ability to effectively utilize these components is critical for enhancing the overall effectiveness of persuasive writing.

The six-factor structure derived from the EFA aligns closely with theoretical models of persuasive writing that emphasize hierarchical and interdependent competencies (Sanders & Wijk, 1996; Wilson et al., 2017). Specifically, the identified factors – Lexical Proficiency, Cohesive Devices, Textual Coherence, Semantic Richness, Structural Effectiveness, and Refutation Quality – reflect the multidimensional nature of writing posited by cognitive ad sociocognitive frameworks, where mastery evolves from foundation linguistic skills (or microstructural elements) to advanced rhetorical strategies (or macrostructural elements). For example, the first three derived factors (i.e., Lexical Proficiency, Cohesive Devices, Textual Coherence) directly map onto the three-level language framework, representing word-level features (e.g., vocabulary diversity, polysemy), sentence-level syntactic cohesion (e.g., logical connectives, referential ties), and discourse-level thematic continuity (e.g., LSA-based coherence metrics). The prominence of Lexical Proficiency (24% variance explained) as a primary factor corroborates theoretical claims that vocabulary depth and complexity are foundational to persuasive efficacy, as lexical choices shape both local clarity and global argumentativeness (Maamuujav, 2022; MacArthur et al., 2019).

This structure also suggests developmental trajectories in writing proficiency. Early writers may rely heavily on Structural Effectiveness and Cohesive Devices to scaffold foundational argumentation, while advanced writers increasingly emphasize Textual Coherence and Refutation Quality to refine rhetorical nuance and audience engagement (L. L. Aull & Lancaster, 2014; Rowe & Wilson, 2015). For example, younger students might focus on formulaic templates (e.g., thesis statements, evidence placement) to meet structural expectations, whereas older learners integrate Refutation Quality to persuade through critical engagement with opposing views. This progression mirrors stage-based theories of writing development (Flower

116

& Hayes, 1981; Scardamalia & Bereiter, 1987; Scardamalia & Paris, 1985), where novices prioritize "knowledge-telling" and experts shift toward "knowledge-transforming".

Overall, the results derived from the EFA provide a productive framework of writing assessment for guiding instructional diagnosis and feedback in a straightforward manner (Berninger et al., 1994; N. W. Nelson & Van Meter, 2007; Scott, 2009). This framework not only facilitates the identification of key components that contribute to effective writing but also allows educators to tailor their feedback to address specific areas of improvement. By focusing on these elements, instructors can enhance their teaching strategies, ultimately fostering better writing skills among students.

Statistically, the EFA revealed six empirically grounded dimensions reflecting the multifaceted nature of persuasive writing. Subsequently, CFA tested a theoretically informed, redesigned model aligned with established argumentative frameworks (Newell et al., 2011). The CFA model refined the structure by: (1) combining Structural Effectiveness and Refutation Quality into a unified construct, reflecting their shared role in the organizational architecture of persuasion; (2) introducing Tone and Content as two separate factors, aligned with their theoretical recognition as critical proxies for persuasive efficacy; and (3) simplifying four EFA-derived microstructural dimensions (Lexical Proficiency, Cohesive Devices, Textual Coherence, and Semantic Richness) into three parsimonious categories (word-level, cohesion, and discourse-level features) to optimize theoretical clarity and operational coherence.

Table 4-3 Factor Structure Coefficients for Micro- and Macro-structural Writing Features

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Communality	Uniqueness	Complexity
DESWLsy	**-0.76**	-0.08	-0.01	0.14	0.32	-0.06	0.702	0.3	1.5
LDMTLD	**-0.52**	0.03	-0.48	0.02	0.09	0.06	0.515	0.49	2.1
PCCNCz	-0.2	0.25	0.08	**0.77**	-0.18	0.11	0.748	0.25	1.6
WRDFAMc	**0.66**	0.17	0.04	0.15	0.09	0.01	0.495	0.5	1.3
WRDFRQc	**0.84**	0.1	0.09	-0.25	-0.11	0.02	0.793	0.21	1.3
WRDHYPnv	-0.42	-0.1	0.1	**0.51**	0.17	0.07	0.49	0.51	2.4
WRDIMGc	-0.35	-0.18	0.02	**0.65**	-0.11	0.13	0.607	0.39	1.9
WRDMEAc	0.1	0	-0.07	**0.83**	0.14	0.01	0.719	0.28	1.1
WRDPOLc	**0.63**	0.08	0.1	0.02	-0.02	-0.03	0.416	0.58	1.1
DESSL	-0.1	*0.46*	0.21	0.12	-0.22	0.07	0.335	0.67	2.2
DESSLd	0.02	*0.43*	0.18	0.01	-0.25	0.07	0.286	0.71	2.1
SYNNP	**-0.67**	-0.07	-0.14	0.19	0.08	-0.03	0.517	0.48	1.3
SYNSTRUTt	-0.03	*-0.48*	-0.14	0	0.15	-0.12	0.289	0.71	1.5
DRPVAL	*-0.28*	-0.1	0	0	0.17	0.07	0.119	0.88	2.1
DRNP	-0.2	*-0.28*	0.1	-0.01	-0.24	0	0.185	0.81	3
DRVP	**0.65**	0.09	0.07	0.04	0.17	0.01	0.47	0.53	1.2
DRNEG	0.3	0.07	-0.01	*-0.32*	-0.01	0.12	0.213	0.79	2.4
PCREFz	0.32	0.21	**0.84**	0.09	-0.03	0	0.855	0.14	1.5
WRDPRP1s	*0.4*	0.05	-0.11	-0.09	-0.1	0.02	0.193	0.81	1.5
WRDPRP2	*0.39*	0.11	-0.05	0	-0.01	-0.13	0.186	0.81	1.4
CRFAOa	0.05	0.27	**0.81**	0.06	0.03	0.03	0.739	0.26	1.2
CRFAO1	0.09	0.23	**0.76**	0.09	0.07	0.02	0.653	0.35	1.3
LSAPP1	-0.17	-0.12	*0.44*	-0.03	*0.44*	-0.01	0.431	0.57	2.5
LSASS1	-0.04	0.12	**0.84**	0.08	0.04	-0.08	0.731	0.27	1.1
LSAGN	0.11	-0.13	**0.66**	-0.06	0.4	-0.09	0.638	0.36	1.8
PCNARz	**0.76**	0.26	0.23	-0.24	-0.15	0.05	0.773	0.23	1.8
PCDCz	0.33	**0.65**	0.17	0.04	0.14	-0.14	0.608	0.39	1.9
DESWC	-0.13	0.06	0.14	0.04	**0.78**	0.15	0.669	0.33	1.2
CNCAll	0.28	**0.82**	0.02	0.05	0.14	-0.11	0.779	0.22	1.3
CNCCaus	0.29	*0.39*	0.17	0.06	0.18	-0.16	0.325	0.67	3.3
CNCLogic	0.34	**0.68**	0.03	-0.12	0.01	-0.03	0.598	0.4	1.5
CNCADC	-0.06	*0.48*	-0.13	-0.25	-0.1	0.19	0.358	0.64	2.2
CNCAdd	0.04	**0.59**	-0.06	0.03	-0.03	0.03	0.352	0.65	1
CNCPos	0.34	**0.62**	0.1	0.12	0.19	-0.21	0.6	0.4	2.2
CNCNeg	0.04	**0.51**	-0.15	-0.18	-0.12	0.2	0.375	0.63	1.9
RDFRE	**0.64**	-0.38	-0.25	-0.17	-0.05	-0.03	0.648	0.35	2.2
Lead	-0.13	-0.09	0.03	-0.04	*0.43*	0.07	0.216	0.78	1.4

Table 4-3 (cont'd)

Position	0.19	0.03	-0.02	-0.02	*0.28*	0.04	0.116	0.88	1.9
Claim	0.07	-0.04	0.12	0.11	**0.55**	-0.05	0.342	0.66	1.2
Counterclaim	0.05	0.06	0.02	0.01	0.2	**0.77**	0.637	0.36	1.2
Rebuttal	0	0.03	-0.01	0	0.24	**0.81**	0.71	0.29	1.2
Evidence	0.04	-0.07	0.02	0.03	*0.4*	0.08	0.17	0.83	1.2
Concluding_Statement	0.01	-0.09	0.01	-0.05	*0.44*	0.05	0.204	0.8	1.1
Lead_effective	-0.18	-0.04	0.01	0.01	**0.62**	0.08	0.431	0.57	1.2
Position_effective	-0.01	0.02	-0.04	0.04	**0.57**	0.06	0.33	0.67	1
Claim_effective	-0.02	0.04	0.02	0.1	**0.75**	0.1	0.582	0.42	1.1
Counterclaim_effective	-0.04	0.06	0.02	0.07	0.33	**0.78**	0.723	0.28	1.4
Rebuttal_effective	-0.07	0.03	0.01	0.05	0.33	**0.75**	0.675	0.32	1.4
Evidence_effective	-0.07	0.01	0.02	0.1	**0.77**	0.1	0.623	0.38	1.1
Concluding_Statement_effective	-0.15	-0.04	0.04	0.05	**0.74**	0.06	0.581	0.42	1.1
C1	-0.09	-0.14	*0.27*	-0.13	-0.1	0.04	0.127	0.87	2.7
C2	0.01	-0.09	*0.45*	-0.06	0.3	0.11	0.314	0.69	2
C6	-0.22	-0.16	*0.44*	-0.03	0.32	0.21	0.408	0.59	3.3
all_markers	0.05	0.12	0.09	0	**0.73**	0.11	0.57	0.43	1.1
booster_words	-0.06	0.05	0	-0.08	*0.48*	0.07	0.25	0.75	1.2
discourse_markers	-0.24	0.11	0.03	-0.07	**0.57**	0.21	0.444	0.56	1.8
negative_adjectives_component	-0.05	0.02	-0.02	-0.14	*-0.2*	0.05	0.064	0.94	2.1
social_order_component	*0.31*	0.01	0.06	-0.04	0.02	0	0.104	0.9	1.1
positive_adjectives_component	*0.4*	0.12	-0.09	0.22	0.19	-0.02	0.271	0.73	2.4
joy_component	*0.34*	0.05	-0.12	0.12	-0.15	-0.02	0.169	0.83	2
trust_verbs_component	*-0.36*	-0.03	-0.03	0.26	0.29	0.01	0.281	0.72	2.8

Note. Factor loadings indicate the strength of the relationship between variables and factors. Communality represents the proportion of variance in a variable explained by the extracted factors. Uniqueness reflects the variance in a variable not explained by common factors. Complexity indicates how many factors are associated with each variable. Bolded values in black indicate the highest absolute factor loadings across the three factors, while italic values indicate absolute factor loadings below 0.50, which have been omitted.

4.4.2 RQ1b. Latent Structure of Persuasive Writing Data

As discussed in Chapter 2 Literature Review, various writing frameworks, such as the

(Not) Simple View of Writing and Levels of Language Framework, represented a spectrum of

multidimensional elements. However, these models often incorporate writing-related elements

that have been insufficiently investigated within realistic educational writing datasets,

particularly those specializing in persuasive writing data (Rodgers et al., 2020). To address the

second part of my RQ1, I employed CFA to establish evidence of construct validity within the

hypothesized structural equation model. I hypothesized that a higher-order model comprising six

first-order factors – word, cohesion, discourse, content, structure, tone/style – would be the most

appropriate representation of the data, explained by both microstructural and macrostructural

factors. It is noted that my hypothesized CFA model differs slightly from the findings of the EFA

presented in RQ1a; nevertheless, I chose to utilize my hypothesis as it is more empirically

grounded and will inform the subsequent design of the GPT prompts and automated essay

scoring system.

I compared this higher-order model to two alternative models: a one-factor model as the

baseline CFA model and a two-factor CFA model. Model fit statistics for each model are

presented in Table 4-4. It is important to note that a reduced version of the higher-order model

was utilized because the full model did not converge. The non-convergence may be attributed to

the ratio of observations to the number of parameters, as well as ambitious model specifications

involving complex paths and correlated factors. Consequently, I opted to simplify higher-order

models by excluding the effectiveness scores of each element at the structural level and

SEANCE sentimental composite scores. For consistency in model comparison, these variables

were also removed from the alternative models to ensure comparable numbers of features across models.

The rationale for removing these variables from all models is twofold. First, these variables are either correlated with or derived from other variables, which could lead to multicollinearity issues. Such multicollinearity can inflate the standard errors of estimates and obscure the true relationships between writing constructs. Second, statistically, the inclusion of highly correlated variables may result in overfitting, which may complicate the model without significantly improving its explanatory power. By simplifying the model and focusing on primary variables that directly contribute to the constructs of interest, I aim to improve model stability and interpretability. Therefore, excluding these variables aligns with both theoretical considerations and the need for a more parsimonious model that can effectively converge during estimation.

Table 4-4 Comparison of CFA model fit indices

Fit Index	One-Factor	Two-Factor	Higher-Order
Free parameters	60	61	67
Chi-Square	43527.447	72162.619	12261.106
Degree of Freedom (df)	405	404	398
p-value	p < 0.001	p < 0.001	p < 0.001
CFI	0.399	0.528	0.853
TLI	0.354	0.492	0.839
RMSEA	0.194	0.172	0.103
SRMR	0.191	0.162	0.107

Note. CFI = Comparative fit index; TLI = Tucker-Lewis index; RMSEA = Root mean square error of approximation; SRMR = Standardized root mean square residual.

The goodness-of-fit statistics for the one-factor model (baseline model) were statistically significant, $\chi^2(405) = 43527.447$, $p < .001$, with both the CFI and TLI indicating poor fit at 0.399 and 0.354, respectively. Similarly, the RMSEA and SRMR were low at 0.194 and 0.191 respectively, neither of which fell below the desirable cutoff of 0.05, indicating poor model fit.

The two-factor model yielded marginal improvements over the one-factor model. The goodness-of-fit statistics were statistically significant, $\chi^2(404) = 72162.619$, $p < .001$, with CFI and TLI values increasing slightly to 0.528 and 0.492, respectively. Additionally, the RMSEA and SRMR values were lower for the two-factor model compared to the one-factor model.

The higher-order model showed a marked improvement in fit indices compared to the previous two models, as evidenced by a significant $\chi^2(398) = 12261.106$, $p < .001$. The CFI and TLI values substantially increased, both exceeding the 0.80 level, although remaining below the 0.90 threshold, which is considered as the minimally acceptable threshold in psychometric research. The RMSEA and SRMR values were 0.103 and 0.107, respectively. Although these values still indicated an undesirable fit, they were approaching the desired cutoff of 0.05. Consequently, I argue that the higher-order CFA model was the most suitable based on these criteria. Statistically, the higher-order model outperformed the other models across all fit indices (i.e., chi-square, CFI, TLI, RMSEA, SRMR).

The weak fit indices for the hypothesized higher-order model (e.g., CFI = 0.853, RMSEA = 0.103) suggest potential misalignment between the theoretical framework and empirical data, likely due to the complexity of modeling higher-order latent constructs. This constrained the robustness of invariance testing for the RQ2, as poor baseline model fit undermines the validity of group comparisons. Future work should prioritize refining the measurement model (e.g., testing bifactor structures or simplifying hierarchical assumptions) to strengthen psychometric foundations before advancing invariance claims. In this study, the hypothesized higher-order structure was retained for theoretical consistency of higher-order structure, despite its constraints on invariance testing.

The standardized loadings for the higher-order CFA model are provided in Table 4-5.

Figure 4-5 illustrates the loading estimates for the higher-order model including both

microstructural and macrostructural writing features.

Table 4-5 Loading Estimate, Standard Error, Z-Value, and P-Value for the Higher-Order CFA Model

Loadings	Estimate	SE	z-value	P(>|z|)
word =~				
WRDFAMc	1			
PCNARz	1.532	0.049	31.048	0.000
DESWLsy	-1.187	0.041	-28.688	0.000
RDFRE	0.621	0.041	15.196	0.000
SYNNP	-1.246	0.048	-26.236	0.000
WRDHYPnv	-0.785	0.038	-20.793	0.000
WRDPOLc	0.990	0.038	25.847	0.000
DRVP	1.054	0.037	28.857	0.000
WRDIMGc	-0.728	0.038	-18.997	0.000
sentence =~				
LSAGN	1			0.000
LSASS1	1.211	0.042	28.838	0.000
PCREFz	1.527	0.063	24.048	0.000
CRFAOa	1.419	0.052	27.411	0.000
CRFAO1	1.372	0.051	26.972	0.000
LSAPP1	0.505	0027	18.426	0.000
text =~				
CNCAll	1			
CNCAdd	0.484	0.021	23.238	0.000
CNCLogic	0.886	0.023	38.213	0.000
DESSL	0.251	0.025	9.866	0.000
PCDCz	0.949	0.026	36.855	0.000
SYNSTRUTt	-0.390	0.025	-15.314	0.000
CNCPos	0.956	0.015	63.440	0.000
content =~				
C1	1			
C2	1.759	0.063	28.041	0.000
C3	1.884	0.065	29.151	0.000
C4	1.908	0.068	28.031	0.000
C5	1.896	0.068	27.696	0.000
C6	1.713	0.065	26.462	0.000
structure =~				
Lead	1			
Position	0.614	0.084	7.356	0.000
Claim	1.531	0.107	14.259	0.000
Counterclaim	0.878	0.082	10.737	0.000
Rebuttal	0.916	0.082	11.144	0.000
Evidence	1.400	0.108	13.000	0.000
Concluding_Statement	1.143	0.085	13.386	0.000
tone =~				
all_markers	1			
booster_words	0.607	0.027	22.661	0.000
discourse_markers	0.744	0.028	26.353	0.000
micro =~				
word	1			
sentence	0.733	0.066	11.112	0.000
text	2.160	0.247	8.731	0.000
macro =~				

Table 4-5 (cont'd)

	Estimate	SE	z-value	P(>|z|)
content	1			
structure	1.902	0.178	10.698	0.000
tone	2.160	0.240	13.302	0.000
Variances	**Estimate**	**SE**	**z-value**	**P(>|z|)**
word =~				
WRDFAMc	0.620	0.021	30.205	0.000
PCNARz	0.109	0.014	7.891	0.000
DESWLsy	0.465	0.017	27.039	0.000
RDFRE	0.854	0.045	18.847	0.000
SYNNP	0.410	0.019	21.910	0.000
WRDHYPnv	0.766	0.023	33.566	0.000
WRDPOLc	0.628	0.021	29.690	0.000
DRVP	0.578	0.020	28.718	0.000
WRDIMGc	0.799	0.023	34.835	0.000
sentence =~				
LSAGN	0.597	0.032	18.778	0.000
LSASS1	0.410	0.018	23.229	0.000
PCREFz	0.061	0.022	2.740	0.006
CRFAOa	0.190	0.013	14.488	0.000
CRFAO1	0.243	0.014	17.875	0.000
LSAPP1	0.897	0.024	37.257	0.000
text =~				
CNCAll	0.191	0.013	14.331	0.000
CNCAdd	0.811	0.026	31.231	0.000
CNCLogic	0.365	0.022	16.742	0.000
DESSL	0.949	0.250	3.792	0.000
PCDCz	0.271	0.021	12.929	0.000
SYNSTRUTt	0.877	0.031	18.010	0.000
CNCPos	0.261	0.018	14.285	0.000
content =~				
C1	0.727	0.028	26.186	0.000
C2	0.155	0.008	20.266	0.000
C3	0.030	0.003	9.211	0.000
C4	0.005	0.004	1.533	0.125
C5	0.017	0.005	3.328	0.001
C6	0.198	0.024	8.098	0.000
structure =~				
Lead	0.860	0.017	50.426	0.000
Position	0.947	0.083	11.349	0.000
Claim	0.672	0.025	26.311	0.000
Counterclaim	0.892	0.051	17.353	0.000
Rebuttal	0882	0.056	15.857	0.000
Evidence	0.726	0.033	21.761	0.000
Concluding_Statement	0.817	0.027	30.165	0.000
tone =~				
all markers	0.063	0.030	2.084	0.037
booster_words	0.655	0.037	17.588	0.000
discourse_markers	0.481	0.033	14.758	0.000
micro =~				
word	0.266	0.031	8.474	0.000
sentence	0.298	0.075	3.992	0.000
text	-0.062	0.014	-4.482	0.000
macro =~				
content	0.161	0.015	10.643	0.000
structure	0.077	0.012	6.184	0.275
tone	0.355	0.030	11.712	0.000
Covariances	**Estimate**	**SE**	**z-value**	**P(>|z|)**
micro~~macro	0.284	0.061	4.638	0.000
Regressions	**Estimate**	**SE**	**z-value**	**P(>|z|)**

Table 4-5 (cont'd)

micro~essay score	-0.209	-0.205	1.020	0.000
macro~essay score	2.729	0.939	2.906	0.000

Note. SE = standard error
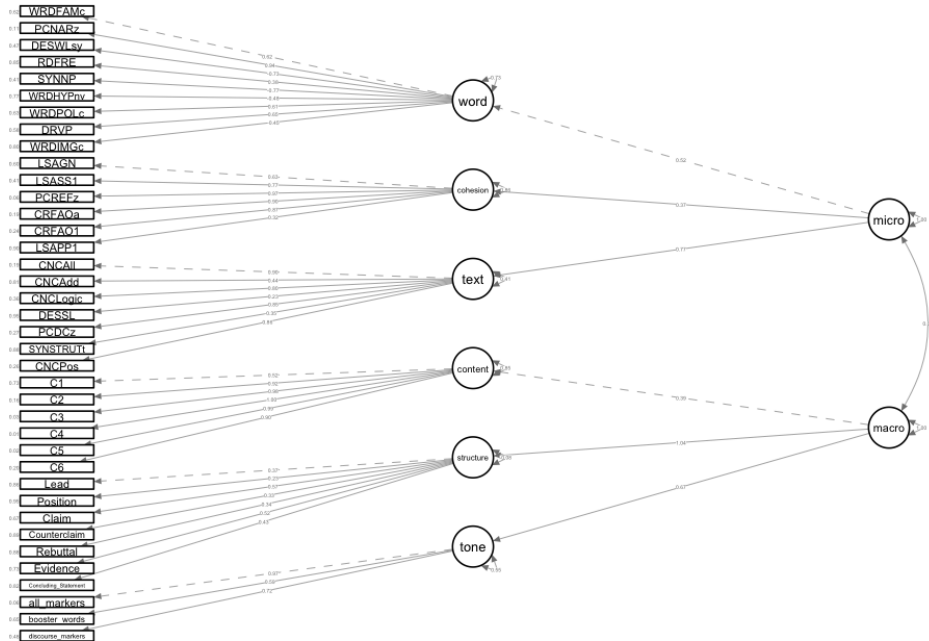
Figure 4-5 Diagram of the Higher-Order CFA Model

4.4.3 RQ2a. Tests of Measurement Invariance Among SPED Groups

The study employed the four-step approach for testing measurement invariance, as outlined by Widaman & Reise (1997), including (1) configural, (2) metric, (3) scalar, and (4) residual invariance. This framework was used to assess whether the resulting reduced higher-order CFA model could detect differences in persuasive writing performance among students with different special education (SPED) status. Table 4-6 presents the fit indices for the models testing measurement invariance.

The analysis began by examining configural invariance, where the same factor structure was specified across SPED groups while allowing all other parameters to vary. The fit indices for this model were suboptimal, $\chi^2(1246) = 26220.821$, RMSEA = 0.136, CFI = 0.648, AIC = 232549.253, suggesting that students with different SPED status did not uniformly conform to the higher-order structure of persuasive writing constructs. However, the baseline model for configural invariance was established for each subscale.

Next, metric invariance was assessed by constraining the same factor structure and requiring equal factor loadings across groups, while allowing all other parameters to vary. The fit indices for this model still reflected a suboptimal fit, $\chi^2(1281) = 26365.663$, RMSEA = 0.135, CFI = 0.646, AIC = 232724.053. The changes in CFI (less than 0.01) and RMSEA (less than 0.015) across different groups (as suggested by Cheung & Rensvold, 2002) supported the acceptance of metric invariance. This indicates the relationships between the latent variables and the observed indicators are equivalent across SPED groups.

Scalar invariance was then tested by constraining the item intercepts to be equivalent across the two SPED groups. The fit indices are $\chi^2(1310) = 27151.004$, RMSEA = 0.143, CFI = 0.638, AIC = 233625.797. While the changes in CFI and RMSEA across different groups remained within acceptable cutoffs, the decrease in CFI by 0.008 suggests a need for caution. Nonetheless, scalar invariance was considered established.

Finally, residual invariance was evaluated by constraining the item residuals to be equal across the groups. The fit indices, $\chi^2(1346) = 28174.077$, RMSEA = 0.136, CFI = 0.616, AIC = 235719.929, indicated that the changes in CFI (above 0.015) across different groups did not meet the required threshold, suggesting that residual invariance was not fully supported. This indicates

significant differences in item residuals across SPED groups, and thus residual invariance was not achieved.

In conclusion, our analysis offers a novel investigation into the measurement invariance of microstructural and macrostructural writing constructs across the SPED groups, which to our knowledge, has not been previously explored. Although the four tests of measurement invariance did not meet the suggested cutoff values for fit indices, the results provide emerging evidence of construct and discriminant validity. Specifically, the number of factors and the relationships between the latent variables and their indicators remained consistent across SPED groups. This suggests that the higher-order CFA model derived from RQ1b is invariant across groups, indicating that any observed group differences in writing constructs likely reflect actual performance differences rather than variations in how participants interpret the instrument.

Table 4-6 Fit indices for the models testing measurement invariance

	Model	Compared Model	χ^2	df	p	$\triangle\chi^2$ (p)	RMSEA	$\triangle$RMSEA	CFI	$\triangle$CFI	AIC
A	Configural invariance		26220.821	1246	0.000		0.136		0.648		232549.253
B	Metric invariance	B vs. A	26365.663	1281	0.000	144.842 (0.000)	0.135	-0.001	0.646	-0.002	232724.053
C	Scalar invariance	C vs. B	27151.004	1310	0.000	785.341 (0.000)	0.134	-0.001	0.638	-0.008	233625.797
D	Residual invariance	D vs. C	28174.077	1346	0.000	1023.073 (0.000)	0.136	+0.002	0.616	-0.022	235719.929

4.4.4 RQ2b. Differences Among the Two SPED Groups

Table 4-7 presents the values of all measurement parameters, including factor loadings, item intercepts, and error variances, for the two SPED groups. In cases where parameters were invariant across both groups, a single parameter value is reported in the table; conversely, differing parameter values are provided where significant variations across groups were observed. This analysis examined the validity of the common assumption of population homogeneity within the microstructural and macrostructural writing construct model. The factor loadings demonstrated full metric invariance across the SPED groups, indicating that both groups utilized the same metric, with the set of factor loadings remaining fully invariant.

Metric invariance was established, indicating that the factor loadings were consistent across groups, which suggests that the same underlying constructs were being measured in the same way in each SPED group. However, scalar invariance was also assessed, and while item intercepts were generally consistent across groups, a slight decline in model fit (e.g., CFI) was observed when scalar invariance was imposed. This indicates that, while the factors maintain equivalent meanings across groups, there may be minor baseline differences in item responses that slightly affected the overall model fit.

According to Table 4-7, the item intercepts and error variances revealed substantial variability. The presence of differing item intercepts across the SPED groups suggests that, while the overall factor structure remains consistent, the starting points (or baselines) of the relationships between the latent variables and the observed indicators vary. It is noted that all writing-related variables, with the exception of LSASS1 and CRFAOa, exhibited opposite trends, indicating that students receiving special education services had lower average response levels for the same writing variables. This pattern may reflect lower writing capabilities or experiences

among students in this group. This finding aligns with previous research indicating that students with disabilities often encounter challenges in various aspects of writing, including vocabulary use (Bryant et al., 2003; O'Connor et al., 2019), cohesion and coherence (Koutsoftas & Petersen, 2017), content development (Graham & Harris, 2010; Monroe & Troia, 2006), structural organization (Golley, 2015; G. A. Troia, 2002), and tone/style (Reed et al., 2023).

The finding that students receiving special education services demonstrated higher measurement errors in specific indicators (e.g., RDFRE, LSAPP1, DESSL, SYNSTRUTt, position, and concluding statement) implies that these indicators may be less reliable for this student population. This variability may indicate that these indicators do not effectively capture the constructs for students with special education needs, potentially due to factors such as learning differences, instructional practices, or the complexity of the writing tasks.

To sum up, the variability observed in the item intercepts and error variances suggests that, while both education groups adhere to a shared framework for the writing constructs, significant differences exist in how these constructs are manifested in the writing of students with varying special education statuses. This finding underscores the necessity for tailored assessment strategies that account for these differences, ensuring that evaluations are equitable and valid for all students. Recognizing that certain writing indicators may operate differently for students in special education can help educators to provide targeted instruction and support. This understanding can inform curriculum design, intervention strategies, and the development of more effective writing assessment tools that better accommodate the needs of students with diverse learning profiles.

Table 4-7 Invariant and Non-Invariant Factor Loadings, Item Intercepts, and Error Variances in Two SPED Groups

Latent Variable	Item	Factor Loadings		Item Intercepts		Error Variances	
		TAP	SPED	TAP	SPED	TAP	SPED
Word	WRDFAMc	0.446		-0.002	0.007	0.662	0.814
	PCNARz	0.726		-0.015	0.063	0.175	0.294
	DESWLsy	-0.645		0.03	-0.128	0.363	0.367
	RDFRE	0.413		0.02	-0.085	0.635	1.269
	SYNNP	-0.634		0.019	-0.082	0.363	0.488
	WRDHYPnv	-0.456		0.064	-0.27	0.643	0.775
	WRDPOLc	0.46		0.026	-0.11	0.659	0.74
	DRVP	0.497		0.007	-0.028	0.602	0.693
	WRDIMGc	-0.354		0.004	-0.015	0.805	0.821
Cohesion	LSAGN	0.584		0.064	-0.271	0.45	0.997
	LSASS1	0.708		0	0.001	0.311	0.512
	PCREFz	0.77		0.006	-0.024	0.21	0.271
	LDTTRc	-0.499		-0.103	0.435	0.6	0.79
	CRFAOa	0.779		0	0.001	0.185	0.281
	CRFAO1	0.767		0.001	-0.005	0.209	0.31
	LSAPP1	0.359		0.067	-0.285	0.748	1.145
Text	CNCAll	0.872		0.022	-0.093	-0.032	-0.079
	CNCAdd	0.589		0.003	-0.014	0.517	0.552
	CNCLogic	0.605		-0.003	0.012	0.465	0.645
	DESSL	0.121		-0.052	0.221	0.537	2.755
	PCDCz	0.61		0.01	-0.042	0.452	0.693
	SYNSTRUTt	-0.219		0.01	-0.044	0.822	1.401
	CNCPos	0.749		0.021	-0.091	0.218	0.295
Content	C1	0.605		0.007	-0.032	0.522	0.682
	C2	0.852		0.093	-0.392	0.103	0.08
	C3	0.883		0.108	-0.456	0.021	0.024
	C4	0.874		0.116	-0.489	0.037	0.02
	C5	0.828		0.137	-0.581	0.113	0.048
	C6	0.74		0.132	-0.557	0.303	0.122
Structure	Lead	0.14		0.078	-0.329	0.95	0.973
	Position	0.076		0.063	-0.268	0.724	2.037
	Claim	0.099		0.094	-0.397	0.96	0.921
	Counterclaim	0.891		0.058	-0.246	0.184	0.196
	Rebuttal	0.862		0.065	-0.274	0.261	0.094
	Evidence	0.217		0.063	-0.265	0.92	0.998
	Cncldng_Sttmnt	0.146		0.078	-0.33	0.849	1.383
Tone/style	all_markers	0.694		0.094	-0.397	0.287	0.146
	booster_words	0.975		0.134	-0.283	0.298	0.123
	discourse_marker	0.556		0.106	-0.362	0.512	0.530

* TAP = typically achieving peers; SPED = students who receive special education services

4.4.5 RQ3a. Automated Persuasive Writing Scoring Algorithm

After addressing the first two research questions, I gained valuable insights into the writing features most indicative of persuasive writing quality among secondary school students. To investigate RQ3, the first step involves building and validating an automated writing scoring system that can provide machine-generated scores for GPT-revised essays based on the identified writing constructs. Specifically, the six-level writing features identified in RQ1 represent salient characteristics that influence persuasive writing quality. These features were subsequently integrated into a pretrained language model, Bidirectional Encoder Representations from Transformers (BERT), to facilitate the prediction of persuasive essay scores.

There are three primary reasons for employing BERT as a baseline model. First, BERT is a transformer-based natural language processing model that has become the *de facto* industry standard for a variety of downstream tasks, including text classification and score prediction (Cochran et al., 2022). The effectiveness of BERT-based transformers for evaluating student responses has been consistently documented (Poulton & Eliens, 2022; Wulff et al., 2023; Zhu et al., 2022). Second, BERT's architecture generates contextualized word embeddings by training on extensive unlabeled corpora, allowing it to capture task-specific nuances and contextual details effectively (Wang et al., 2024). Third, once the BERT model is successfully trained and validated, future datasets can be processed without the need to reintroduce features from external sources. Different from other supervised machine learning models, the BERT model can directly utilize the embedded contextual information for various tasks.

To develop and validate the automated essay scoring model, I utilized the pre-trained BERT model from the Huggingface library and configured it to align with the parameter of the basic version of Google BERT (Devlin et al., 2019). The model included 12 self-attention layers,

12 attention heads, and a hidden dimension of 768 for the embedding vectors. During fine-tuning, the pre-trained models were concatenated with a text classification model. In the training phase, essays (n = ~25,000) from the PERSUADE corpus were input into BERT to generate feature vectors. The mean of these vectors was forwarded to the text classification layer to generate predictions based on six scoring levels. The cross-entropy loss was then computed by comparing the predictions to the ground truth.

For these experiments, I employed the Adam optimizer with consistent hyperparameters: a learning rate of 1e-04, a batch size of 10, an input sequence length of 128, 20 training epochs, a weight decay of 0.0005, and a dropout rate of 0.3. The training was performed on an NVIDIA 1080Ti GPU. All experiments involved training, testing, and validating the model on original student essay scores, with a test size of 0.3.

In this section, I report key evaluation metrics – MAE, standard deviation of MAE, R^2 – for evaluating and comparing the performance of the two auto-scoring models in predicting persuasive essay scores: (1) the BERT-generic model as the baseline model and (2) the BERT model incorporating the EFA-identified writing features. Performance results for all approaches are presented in Table 4-8.

Table 4-8 Results of Two BERT Models on Scoring Prediction Task

Models	MAE (SD)	R^2
BERT	0.65 (0.67)	0.66
BERT+writing features	**0.42 (0.33)**	**0.82**

Note. MAE = Mean Absolute Error, SD = Standard Deviation

Both BERT models demonstrated strong performance, with BERT model incorporating writing features consistently outperforming the generic BERT model across all evaluation metrics. The BERT pretraining model with EFA-derived writing features achieved higher scores in all metrics, particularly for R^2, indicating that 82% of the variance in persuasive essay scores

133

was explained by this model. This finding suggests a robust alignment between machine-generated scores and human-rated scores, with the model also demonstrating greater stability, evidenced by a lower standard deviation of the MAE. Given the successful performance of this model on the PERSUADE dataset, I have selected it as the final model for further analyses. This model will be used to automatically score GPT-revised essays and compare their performance to the original student essays, evaluating improvements in quality based on key writing features. Figures 4-6 and 4-7 present scatter plots illustrating the predicted and actual results for the BERT-generic model and the BERT model enhanced with writing features, respectively.
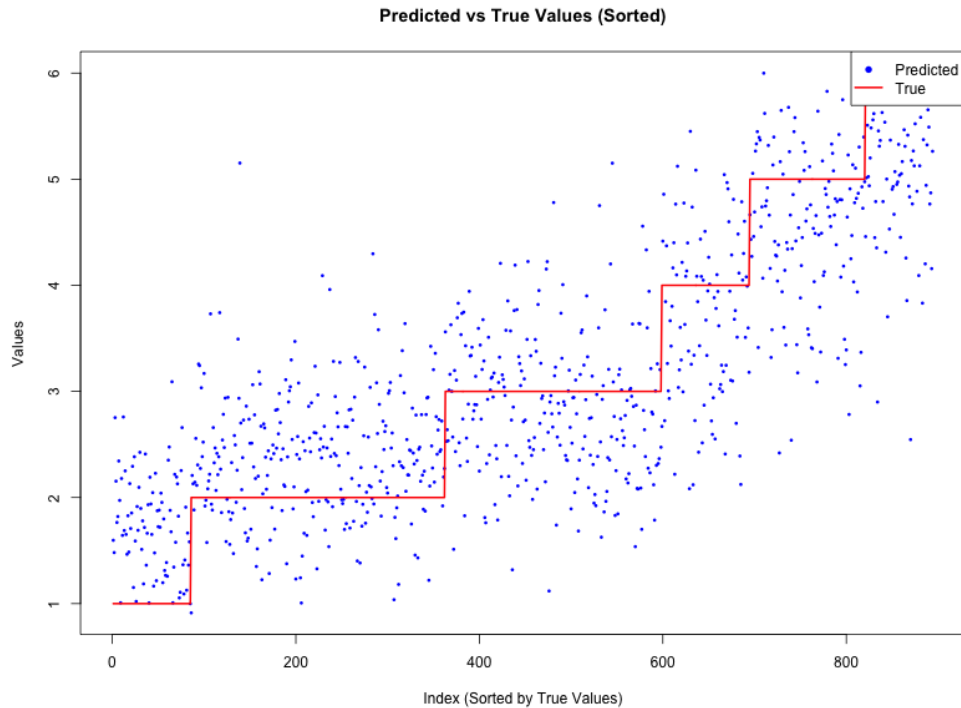
Figure 4-6 Predicted and True Values in BERT-Generic Model

Predicted vs True Values (Sorted)
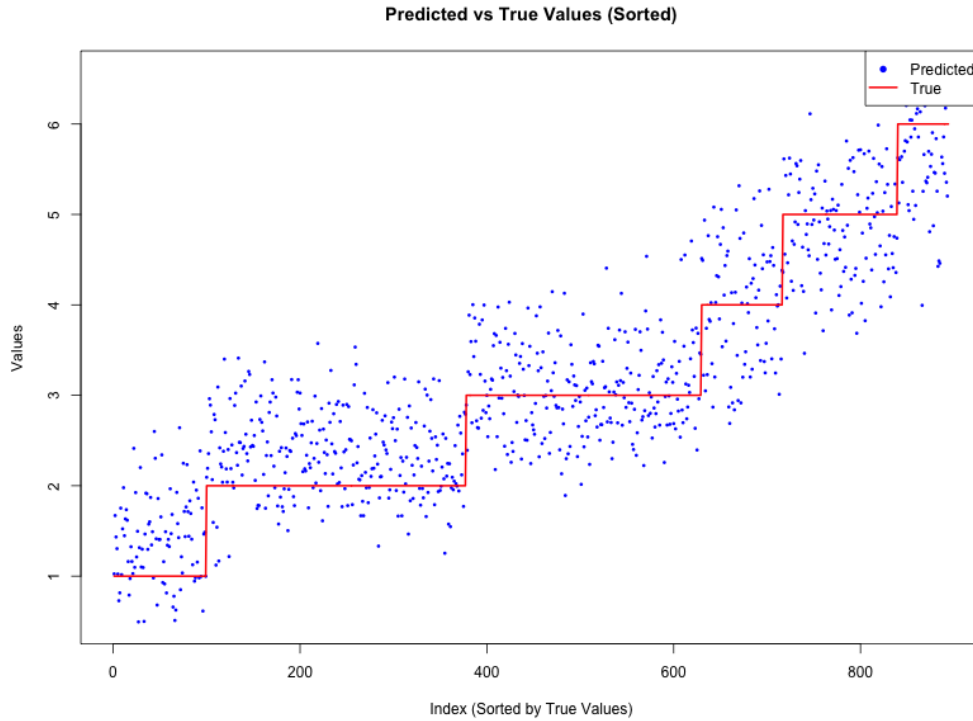
Figure 4-7 Predicted and True Values in BERT Model Enhanced with Writing Features

4.4.6 RQ3b. Effectiveness of GPT-Revised Essays

The prompt for generating feedback (see Table 4-9) using GPT was developed based on the insights gained from the first two RQs and relevant feedback literature (Hattie & Timperley, 2007; Graham & Sandmel, 2011). Drawing from Meyer et al. (2024), and the prompt instructed GPT to provide students with suggestions for improving their text by offering hints and examples. Specifically, building on RQ1 and RQ2 findings, GPT was directed to focus feedback on six key aspects of persuasive essay quality: lexical proficiency, cohesion and coherence, text complexity, structure, content, and tone/style. To minimize cognitive load for students, the feedback was required to be structured, with short examples from the student's own text for individualized guidance. Additionally, the prompt incorporated a revised version of the student essays to assess whether GPT-revised essays could serve as effective mentor texts aligned with the six writing dimensions identified through factor analysis. The GPT settings included the

GPT-3.5-turbo model, a temperature of 0.7, and a maximum length of 1800 (see Figure 4-8).

Examples of the generated feedback are provided in Figure 4-9.

The phrasing used in the GPT prompts aligns with the CFA-derived latent constructs (see

Tables 4–5 and Figure 4–5), which are grounded in the hypothesized theoretical framework.

Minor terminological divergence (e.g., replacing the factor label "word" with the prompt

phrasing "lexical proficiency" or "text" with "discourse complexity") reflects intentional

adjustments to maintain the theoretical integrity of the constructs while optimizing clarity for

generative AI applications.

Table 4-9 Prompt For GPT

Please read the following instruction step by step: In the following you will read a persuasive essay of a secondary student. Give elaborated feedback on the following text on six aspects: (1) lexical proficiency; (2) cohesion and coherence; (3) discourse complexity; (4) content; (5) structural efficiency; and (6) tone/style. Give feedback in a highly structured manner displayed as a table with one column for the six aspects with one row for each aspect (lexical proficiency, cohesion and coherence; discourse complexity; content; structural efficiency; tone/style), one column with hints for improvement on the relevant aspect, and one column with three examples for improvement on the relevant aspect. The examples should include SHORT aids to revise the essay, such as sentence beginnings or transition words. The examples must NOT contain fully formulated sentences. Provide three examples per aspect. Write hints as fully formulated sentences. The suitable examples for the corresponding aspect should be presented as key points. Use the following table structure as an example and use the table content as orientation for your own feedback:

```
<table class="MsoTableGrid" style="border-collapse:collapse; border:none; line-height: 1.5;" cellpadding="10" >
    <tbody>
        <tr>
            <th style="width: 15%; border:1px solid black;" valign="top">Aspect</th>
            <th style="width: 50%; border:1px solid black;" valign="top">Hints for improvement</th>
            <th style="border:1px solid black;" valign="top">Examples for improvement</th>
        </tr>
        <tr>
            <td style="width: 15%; border:1px solid black;" valign="top"> lexical proficiency </td>
            <td style="width: 35%; border:1px solid black;" valign="top">hint 1 for lexical proficiency/hint 2 for lexical proficiency/hint 3 for lexical proficiency</td>
            <td style="border:1px solid black;" valign="top">example lexical proficiency 1 / example lexical proficiency 2 / example lexical proficiency 3</td>
        </tr>
        <tr>
            <td style="width: 15%; border:1px solid black;" valign="top"> cohesion and coherence </td>
            <td style="width: 35%; border:1px solid black;" valign="top"> hint 1 for cohesion and coherence / hint 2 for cohesion and coherence /hint 3 for cohesion and coherence </td>
            <td style="border:1px solid black;" valign="top">example cohesion and coherence 1 / example cohesion and coherence 2 / example cohesion and coherence 3</td>
        </tr>
        <tr>
            <td style="width: 15%; border:1px solid black;" valign="top">discourse complexity</td>
```

Table 4-9 (cont'd)

                    \<td style="width: 35%; border:1px solid black;" valign="top">hint 1 for language/hint 2 for discourse complexity/hint 3 for discourse complexity\</td>

                    \<td style="border:1px solid black;" valign="top">example discourse complexity 1 / example discourse complexity 2 / example discourse complexity 3\</td>

          \</tr>

          \<tr>

                    \<td style="width: 15%; border:1px solid black;" valign="top"> content \</td>

                    \<td style="width: 35%; border:1px solid black;" valign="top">hint 1 for content/hint 2 for content/hint 3 for content \</td>

                    \<td style="border:1px solid black;" valign="top">example content 1 / example content 2 / example content 3\</td>

          \</tr>

          \<tr>

                    \<td style="width: 15%; border:1px solid black;" valign="top"> structural efficiency \</td>

                    \<td style="width: 35%; border:1px solid black;" valign="top">hint 1 for structural efficiency/hint 2 for structural efficiency/hint 3 for structural efficiency\</td>

                    \<td style="border:1px solid black;" valign="top">example structural efficiency 1 / example structural efficiency 2 / example structural efficiency 3\</td>

          \</tr>

          \<tr>

                    \<td style="width: 15%; border:1px solid black;" valign="top"> tone/style \</td>

                    \<td style="width: 35%; border:1px solid black;" valign="top">hint 1 for tone and style/hint 2 for tone and style/hint 3 for tone and style\</td>

                    \<td style="border:1px solid black;" valign="top">example tone and style 1 / example tone and style 2 / example tone and style 3\</td>

          \</tr>

     \</tbody>

\</table

In addition, based on the feedback provided, please revise this student's persuasive essay according to the six writing aspects.

Here is the essay: {student_essay}

```
# Make a request to the GPT model (you can adjust the model and parameters as needed)
response = openai.ChatCompletion.create(
    model="gpt-3.5-turbo",  # You can use "gpt-3.5-turbo" as well
    messages=[
        {"role": "system", "content": "You are a writing expert who provides structured feedback."},
        {"role": "user", "content": prompt}
    ],
    temperature=0.7,
    max_tokens=1500  # Adjust depending on how long you expect the response to be
)

# Extract and print the feedback from the response
feedback = response['choices'][0]['message']['content']
```

Figure 4-8 Settings for GPT API Feedback Generation

Figure 4-9 Example 1: A 6th-grade student who is not identified as having disability received a score of 3 for the essay (revised essay score = 4.55)

Aspect	Hints for Improvement	Examples for Improvement
Lexical Proficiency	1. Use precise and varied vocabulary to make the argument more compelling and informative. 2. Replace general or repetitive words with more specific terms to improve clarity. 3. Avoid informal language to maintain a formal tone suitable for persuasive writing.	*Use "fosters creativity" instead of "bring out creativity" | Replace "better your work" with "enhances the quality of your work" | Use "select" instead of "pick"*
Cohesion and Coherence	1. Add transitional phrases to clarify the connection between ideas, especially between paragraphs. 2. Introduce topic sentences that summarize each paragraph's main argument for improved flow. 3. Use concluding phrases at the end of each paragraph to reinforce your main point.	*"Furthermore" | "In addition to this" | "As a result, teacher-designed projects help students…"*
Discourse Complexity	1. Vary sentence structures by using a mix of simple, compound, and complex sentences to enrich the argument. 2. Avoid run-on sentences by using conjunctions appropriately. 3. Combine related ideas into single complex sentences where possible for a more sophisticated tone.	*"Teacher-designed projects, which provide…" | "Students benefit from…" | "Because these projects prepare students, they are…"*
Content	1. Support arguments with specific examples or evidence, such as the benefits of learning continuity in teacher-designed projects. 2. Expand on the counterarguments by explaining why student-designed projects may seem appealing but lack the same educational value. 3. Provide a clear stance by emphasizing why teacher-designed projects provide long-term benefits.	*"Research shows that…" | "For instance, projects focusing on…" | "Teacher-designed projects align with upcoming curriculum, unlike student-chosen topics…"*
Structural Efficiency	1. Organize arguments into clear, distinct sections with a single focus per paragraph for stronger structure. 2. Ensure each paragraph has a topic sentence, body, and concluding sentence to support clarity. 3. Consolidate similar points into one paragraph to avoid redundancy.	*"In summary…" | "The main reason for…" | "An additional benefit is…"*
Tone/Style	1. Maintain a formal tone by avoiding contractions and conversational expressions. 2. Use confident language to reinforce the persuasiveness of the argument. 3. Replace casual language with academic expressions where possible.	*Use "it is essential to…" instead of "it's easier…" | "demonstrates that…" instead of "shows how…" | "contribute to learning" instead of "help with learning"*

Figure 4-9 (cont'd)

Revised Essay

Some schools require students to complete summer projects to ensure they continue learning during their break. There is an ongoing debate about whether these summer projects should be teacher-designed or student-designed. Some argue that student-designed projects are easier and encourage individual creativity; however, this view overlooks significant educational benefits. Although some may disagree, summer projects should be teacher-designed.

There are several reasons why summer projects should be teacher-designed. Firstly, teacher-designed projects challenge students in ways that self-selected topics often do not. For example, if students choose their own topic, they may not choose a task that encourages rigorous thinking. Additionally, a teacher-designed project can prepare students for the upcoming year by focusing on topics they will soon study. Completing a project of my own choice, on the other hand, may not support my academic growth for next year. Furthermore, teacher-designed projects allow teachers to evaluate student skills, providing insights into who may need additional help. Each school year is different, and completing a structured summer project can help students start the year prepared.

There are additional reasons why summer projects should be teacher-designed. First, these projects are aligned with the subjects students will study. Doing a project about a favorite song, for example, would not contribute to learning in science. By designing projects around academic topics, teachers help students strengthen their skills and improve the quality of their work. Moreover, summer projects prepare students' minds for school. During the summer, students may forget some of what they learned, but a teacher-designed project helps them retain important skills and knowledge.

Although some may disagree, summer projects should be teacher-designed due to their benefits in improving work quality, preparing students for the next year, and challenging students academically. Those who disagree argue that student-designed projects are easier and allow for creativity, but this view overlooks the learning advantages of teacher-designed tasks. This topic is important because it highlights how teacher-led summer projects support students' educational readiness, encouraging other schools to adopt similar approaches to prepare students for the upcoming year.

Figure 4-10 Example 2: A 11th-grade student who is not identified as having disability received a score of 6 for the essay (revised essay score = 5.79)

Aspect	Hints for Improvement	Examples for Improvement
Lexical Proficiency	Consider using more precise vocabulary to avoid repetition and create a more professional tone. Replace colloquial expressions with academic language to enhance persuasiveness. Aim for varied adjectives to better convey meaning.	- Use "critical" instead of "important" - Replace "lots of" with "numerous" - Use "academic break" instead of "long break"
Cohesion and Coherence	Improve the flow by using transition words to connect sentences and paragraphs more smoothly. Avoid sudden topic shifts by linking ideas logically. Clarify any ambiguous points for better readability.	- Begin sentences with "Consequently," "Moreover," or "In addition" - Use transitions like "for instance" or "specifically"
Discourse Complexity	Vary sentence structures to enhance readability and avoid redundancy. Combine short sentences where possible and use complex sentences to convey nuanced ideas. Vary sentence length for reader engagement.	- Combine sentences like "This means..." - Use "although" to contrast ideas - Start sentences with phrases like "Given that..."
Content	Add more specific details to support each argument, with examples or evidence to strengthen claims. Address counterarguments to provide a balanced perspective. Expand points with illustrative details.	- Add examples like "For example, student-designed projects..." - Use comparisons like "Unlike teacher-designed assignments..."
Structural Efficiency	Organize the essay clearly, introducing main points early and concluding each paragraph effectively. Ensure each paragraph flows logically to the next, maintaining a coherent argument throughout.	- Start paragraphs with "Firstly," "Secondly," or "Finally" - Conclude with "This demonstrates that..." - Begin conclusion with "In summary"
Tone and Style	Use a formal tone by avoiding conversational language. Aim for an authoritative style that reflects confidence. Maintain consistency in tone to strengthen the argument's impact.	- Use "students may" instead of "students might" - Replace "kids who would" with "students who tend to" - Avoid contractions

Figure 4-10 (cont'd)

Should Summer Projects Be Student-Designed?

Reflecting on elementary school, many students remember being assigned summer reading logs and lists of books to complete during their break. While some students eagerly embraced these assignments as an opportunity to read, others found them challenging or reluctantly completed them. In elementary school, whether students completed the reading logs did not significantly impact their grades. However, as students reach high school, summer projects become actual requirements. While some students argue that it is unfair to assign work over the one long break they receive, summer assignments can be beneficial if structured effectively. Schools should consider making these projects student-designed rather than teacher-designed. Student-designed projects would not only encourage greater engagement but would also allow students to balance their summer plans more effectively.

Summer represents the only substantial break that students receive from schoolwork, assignments, and tests. From August to June, school is a constant commitment, with students following a structured schedule five days a week. This rigorous schedule means that students need time off to rest and recharge. Summer provides an ideal opportunity to relax, spend time with friends and family, travel, and enjoy new experiences. Assigning a lengthy, teacher-designed project over the summer can interfere with these plans. Imagine a student who, on the last day of school, receives a project requiring an eight-page essay on the use of propaganda in society. Such a task would likely seem overwhelming and even discouraging. By contrast, if the project were student-designed, students could create work on topics of interest to them, which could inspire more engagement and creativity. For instance, students could explore topics they find intriguing, like endangered species or environmental challenges, making the project feel less like a chore and more like an opportunity for personal expression.

Summer is also a time for family vacations and trips, both locally and internationally. Each trip requires planning and preparation, often limiting the time and resources students have available for academic projects. In my own experience, I once spent an entire summer traveling, from visiting relatives in Pennsylvania to swimming in the clear waters of the Bahamas. Completing a complex teacher-designed assignment during that time would have been nearly impossible. Students need the flexibility to design projects that can fit around their travel schedules. For example, if students are tasked with researching a local topic, they may need access to public libraries, which isn't possible if they are on vacation. Student-designed projects allow students to plan their work around their summer activities, determining when they will have time and access to the resources they need.

Some educators may argue that giving students control over their summer projects could lead to minimal effort and lower-quality work, as students might prioritize leisure over learning. However, because these projects are graded, students are unlikely to put in minimal effort if it will negatively affect their grades. The potential impact on their academic record would motivate them to produce quality work, even if they have some freedom in designing the project. The project's overall benefits would encourage students to invest time and effort, knowing that a thoughtful, well-executed project would support their success in the coming school year.

In conclusion, student-designed projects offer unique advantages for both teachers and students. By allowing students to choose topics they find meaningful, these projects encourage creativity, personal expression, and increased motivation. Teachers, in turn, benefit by gaining insight into their students' interests, which may help them better connect with their students throughout the school year. Requiring teacher-designed projects during summer break risks burdening students and detracting from their well-deserved rest, while student-designed projects could transform the experience into one of growth, learning, and enjoyment.

Figure 4-11 Example 3: A 10th-grade student who is identified as having disability received a score of 2 for the essay (revised essay score = 4.52)

Aspect	Hints for Improvement	Examples for Improvement
Lexical Proficiency	1. Avoid informal language; try to use more precise vocabulary. 2. Replace vague terms like "cool" and "crazy" with words that better convey your stance. 3. Ensure correct spelling, especially for commonly misspelled words like "especially" and "business."	"interesting" instead of "cool" / "controversial" instead of "crazy" / "public use" instead of "places for crazy people"
Cohesion and Coherence	1. Improve transitions between ideas to clarify your argument's flow. 2. Separate long, run-on sentences into shorter ones for better readability. 3. Use connecting phrases to emphasize the reasoning behind each point.	"Although..." / "In contrast, teachers might..." / "Additionally, this technology could..."
Discourse Complexity	1. Develop your points in depth to make your argument more persuasive. 2. Incorporate examples and specific scenarios to illustrate each argument clearly. 3. Use conditional phrases to show hypothetical situations (e.g., "If...then...").	"If used in schools, teachers might..." / "For instance, if a student..." / "This technology might lead to..."
Content	1. State your main argument at the beginning for clarity. 2. Provide reasons for your stance in separate sentences to strengthen the argument. 3. Add real-world implications or specific examples to support your perspective on privacy concerns.	"The device could infringe on privacy..." / "Students might feel uncomfortable knowing..." / "In a school setting, this could..."
Structural Efficiency	1. Follow a clear structure: introduction, supporting arguments, and conclusion. 2. Avoid repeating phrases (e.g., "to sum this all up in my conclusion") to create a more professional tone. 3. Use separate paragraphs for each major idea.	"This essay argues..." / "Firstly..." / "Secondly, another concern is..."
Tone/Style	1. Use a formal, objective tone for a persuasive essay. 2. Avoid conversational fillers like "yeah" or "okay" to maintain a professional tone. 3. Choose words that reflect the topic's importance, emphasizing why it matters for students and schools.	"I strongly believe..." / "It is essential to consider..." / "This issue could have serious effects on..."

144

Figure 4-11 (cont'd)

Revised Essay

The idea of having a piece of technology that could show and reveal people's emotions is fascinating, but also raises several ethical and privacy concerns, especially when applied to school settings.

To begin, I am opposed to using this technology in schools. While it could be interesting to see others' emotions and understand their feelings, there are privacy issues that must be considered. Teachers, for example, might rely on this tool whenever a student is upset or behaving unusually, such as after an argument. With this technology, a teacher could detect a student's feelings, potentially intruding on personal privacy. For instance, if a student is texting quietly or reacting emotionally, a teacher could use this device to determine what the student is feeling. While this might help teachers identify when students are distracted, it could also lead to frustration among students, making them feel uncomfortable or invaded, which could harm their view of school.

In conclusion, while this type of technology could offer unique insights in certain public spaces, it does not belong in schools. Instead, it should be placed in locations where understanding emotions can serve a positive public purpose, such as in hospitals where healthcare professionals could benefit from knowing patients' emotions. However, schools should remain spaces where students' feelings are respected and not monitored by devices.

After obtaining the GPT revised essays, they were then input into the previously built automated essay scoring systems in RQ3a and predicted the essay scores. The original essays yielded a mean score of 4.49, with a standard deviation of 1.00 and a range from 1 to 6. The GPT-revised essays had an increased mean score of 5.33, with a standard deviation of 0.26 and a range from 2.63 to 5.95. A paired t-test revealed a highly significant difference between the original and GPT-revised scores, $t(14998) = 33.21$, $p < 0.001$, indicating a substantial increase in scores following the revisions. The effect size, Cohen's $d = 1.063$ [0.984, 1.142], demonstrates a large effect, reinforcing the effectiveness of the GPT revisions.

The results were aligned with my hypothesis that essays revised by GPT based on the prompts designed for improving students' writing in multidimensional levels can be effective as mentor texts and supporting students during text revision. This investigation contributes to the existing literature by providing empirical evidence of the potential that GPT has for improving student persuasive writing with a tailored and efficient way to help students improve their writing quality better. The distribution of GPT-revised essay scores by original essay score levels, as shown in Figure 4-10, indicates that the majority of GPT-revised essays achieved scores exceeding level 4.

Figure 4-12 Boxplot of GPT Revised Essay Scores by Original Essay Score Levels

Chapter 5: Conclusion

This final chapter presents a comprehensive discussion of the study's implications. It further addresses the limitations encountered throughout the research and explores potential avenues for future research, suggesting how the findings can be extended or applied in different contexts. Lastly, the chapter concludes the thesis by summarizing the key contributions and insights gained from the research.

5.1 IMPLICATIONS

This study employed both psychometric approaches and artificial intelligence techniques to explore how GPT, when guided by prompts designed through human intelligence, can generate revisions for students' original persuasive texts by considering key writing constructs. Specifically, the study addresses three major research questions. The first research question aimed to identify the most representative features that capture the quality of students' persuasive essays. This was achieved through exploratory and confirmatory factor analyses, with the resulting features serving as foundational predictors for establishing an assessment framework based on six writing dimensions. The second research question examined the measurement invariance of these writing features across different student populations, specifically comparing typically achieving students with those receiving special education services, using multigroup confirmatory factor analysis. The third research question explores the integration of human intelligence with machine intelligence by prompting GPT to revise student essays in alignment with the six dimensions of writing constructs that students are expected to achieve in composing persuasive essays. Overall, the findings offer potential avenues for leveraging AI to create low-cost, accessible learning opportunities for students, particularly in content-rich subjects such as

English Language Arts or other disciplines that require writing as a means of activating and applying knowledge.

The findings from this dissertation offer several practical implications for both teaching and assessment. First, regarding writing instruction, the writing features and six constructs identified through exploratory and confirmatory factor analyses can inform instructional decision-making and guide targeted feedback. These features were selected not only on their statistical relevance – where specific writing features were found to significantly impact writing quality – but also on a robust theoretical framework that incorporates both microstructural and macrostructural aspects of writing from surface-level elements to deeper layers of meaning (Aryadoust & Liu, 2015; Siekmann et al., 2022). For example, the factor analysis revealed that lexical proficiency explained 17% of the variance in persuasive writing quality, with WRDFRQc identified as the highest loading lexical feature. This suggests that the use of rich, appropriate, and complex academic vocabulary plays a critical role in persuasive writing, and that a higher frequency of content words is a key indicator of lexical proficiency. As a result, educators are encouraged to provide comprehensive and individualized vocabulary instruction that focuses on improving students' lexical choices in content-rich persuasive essays.

Furthermore, the findings from RQ1 and RQ2, which identified six key writing dimensions underlying large-scale persuasive writing essays (e.g., lexical proficiency, argumentative structure, rhetorical coherence), provide empirically validated insights into the latent dimensions of persuasive writing proficiency. These insights enable educators to deliver targeted, diagnostic feedback to struggling writers. By pinpointing specific factors – such as weak logical sequencing or insufficient use of credible evidence – that underpin a student's writing difficulties, teachers can prioritize interventions that address discrete skill gaps rather

than relying on generalized critiques. For example, if students underperform on the *content* dimension – reflecting gaps in domain-specific understanding or factual accuracy – teachers might prioritize writing instructional practices such as pre-writing research tasks, collaborative fact-checking exercises, or guided source integration to build content knowledge. Deficiencies in the *cohesion and coherence* dimension (e.g., ineffective use of connectives and transitional phrasing, fragmented sentence flow) could warrant interventions like sentence-combining activities or explicit instruction on paragraph-level coherence. For students with disabilities, detecting gaps in cohesion and coherence poses challenges, as they often struggle to independently identify disjunctures in logic or tone. To scaffold this process, revising checklists derived from these dimensions—for instance, prompts like "Does each paragraph advance my central argument?" or "Are ideas linked with appropriate transitional phrases?"—can systematize self-assessment protocols and guide structured peer review during revising and editing stages. Similarly, weaknesses in the *structural efficiency* dimension (e.g., disorganized argumentation, insufficient evidence support, lack of counterarguments) might necessitate interventions, such as scaffolded outlining templates or teachers' modeling of integrating core argumentative components during planning and drafting period. By aligning feedback with factor analysis identified skill gaps, educational practitioners can systematically address the hierarchical and interdependent nature of persuasive writing competencies, fostering growth in both metacognitive awareness and procedural mastery (Negretti, 2012). Second, this study highlights the potential benefits of adopting technology to support student learning. Research indicates that teachers who effectively integrate technology into their teaching can reduce their workload (Gregory & Lodge, 2015; Ming, 2005), enhance student engagement and motivation (Y.-T. Chuang, 2014; Williams & Beam, 2019), and provide more individualized learning opportunities

(Brown & Green, 2014; Fuchs, 2023; Wang et al., 2023), particularly in complex domains such as writing. Given the rapid growth of AI-based technologies and their increasing relevance in the digital age, educational researchers have called for the redesign of technology-mediated writing instruction to better support students' writing development (Özer & Yükselir, 2022; Soltyska, 2024). While a national survey (Graham et al., 2014) noted that middle school teachers rarely used technology as part of their writing programs, more recent surveys and case studies (Bateman, 2024; Turcotte & Mesa, 2024) have shown that teachers are beginning to recognize the benefits of content-generation tools such as ChatGPT in curriculum design, teaching, assessment, and feedback. However, concerns remain about potential biases, discrimination due to reliance on natural language processing, and the risk of students plagiarizing texts without critical analysis (ElSayary, 2024). This study proposes a solution for maximizing the benefits of ChatGPT: teachers can provide GPT with a rubric outlining their expectations for student essays, and GPT can generate texts aligned with those standards. Additionally, constructive feedback comparing the strengths and weaknesses of students' original essays with the expected revisions could foster critical thinking and improve writing skills (Valero Haro et al., 2024). This approach has the potential to support personalized learning for struggling writers and students with learning difficulties.

In practice, teachers can pair their assessment rubrics with AI-generated mentor texts that exemplify proficiency in each dimension (e.g., annotated essays demonstrating effective transitions or evidence integration) by providing students with concrete models of success. For example, a student struggling with cohesion might analyze a mentor text highlighting connective language and then apply a rubric-aligned checklist to revise their own transitions. AI-driven tools can further leverage these dimensions to generate targeted feedback, such as flagging

underdeveloped arguments (weak structural efficiency) or repetitive vocabulary (low lexical proficiency) and recommending specific revisions. This approach shifts writing instruction from generic feedback to skill-specific interventions, which can empower students to address discrete weaknesses and foster self-regulated revision. For learners with disabilities, this granularity can reduce cognitive overload by breaking writing into manageable, rubric-anchored tasks. By aligning instruction, assessment, and AI content generation tools with empirically validated dimensions, teachers can scaffold mastery of complex persuasive writing processes, promoting equity through personalized, competency-based support.

Finally, in terms of assessment, the integration of representative persuasive writing features from factor analysis into a pretrained language model, such as BERT in this study, proved to be an effective strategy of scoring prediction. The results showed that the BERT model incorporating these writing features outperformed the generic BERT model in predicting persuasive writing scores. This enhanced model is well-suited for assessing the quality of student persuasive writing by considering six dimensions: word, cohesion, discourse, content, structure, and tone/style. In practice, this automated scoring system, combined with GPT's feedback generation, can be preserved for future use. As new essays are submitted, these systems can provide instant feedback and revised versions to students, offering tailored support in line with individual learning needs and academic goals.

5.2 LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

This study has several limitations that should be considered when interpreting the findings. First, it is important to note that only one form of generative AI, GPT-3.5, was utilized in this research. Given the rapid advancements in AI-driven language models, future versions may yield different results in terms of essay revisions and feedback. Future research should

evaluate the robustness of a variety of AI-based tools by incorporating a broader range of content generation technologies and examining their impact on writing education on a larger scale.

Second, the feedback prompts used in this study were guided by recent research on AI feedback in writing (Meyer et al., 2024; Steiss et al., 2024) and were aligned with key writing features identified through exploratory and confirmatory factor analysis, which were deemed important for persuasive writing. While my primary goal was to provide the most beneficial feedback by focusing on informative writing constructs, effectively incorporating GPT into classroom settings will likely require teachers to modify curricula and instructional strategies. The success of AI tools in enhancing writing development largely depends on how teachers facilitate student use of these technologies (McKnight et al., 2016; Runge et al., 2023). Future studies could employ experimental designs or survey-based studies to assess the efficacy of AI-generated feedback compared to traditional methods, or by examining student perceptions of using GPT for essay revisions.

Third, AI-generated feedback may not have the same impact across all student demographics. For instance, English Language Learners and students with writing difficulties might require additional support to fully engage with AI-provided feedback (Warschauer et al., 2023; Tate et al., 2024). Future studies should consider developing instructional strategies that consider individual learning needs, personality traits, and educational goals (Panadero & Lipnevich, 2022), while also examining whether AI feedback produces consistent results across different student populations.

Fourth, the selection of algorithms for auto-scoring persuasive writing may have influenced the outcomes of this study. Previous research suggests that transformer-based language models, such as BERT, generally outperform other machine learning models in various

contexts (Ludwig et al., 2021). However, different evaluation metrics can lead to varying results depending on the specific application (Doewes et al., 2023). Consequently, future research should explore the development and validation of customized models for different use cases, being mindful of potential algorithmic biases when replicating results across datasets.

Fifth, the present study did not incorporate certain writing metrics into the design of the automated essay scoring algorithms, leaving an opportunity for future research to build on this work by integrating additional writing-related features. For instance, measures of conceptual density—such as the frequency of Tier 2 and Tier 3 vocabulary within sentences or paragraphs, which can reflect students' writing proficiency—could be explored. Complex sentences, in particular, may include academic or discipline-specific vocabulary used relationally within a single sentence (Englert et al., 2009). Additionally, content knowledge – another foundational element of persuasive writing – was not explicitly investigated in this study. Domain-specific understanding enables writers to construct credible arguments, integrate evidence effectively, and anticipate counterarguments (Ferretti et al., 2009; Wijekumar et al., 2019). While its influence is implicitly reflected in the broader framework of argumentative structure and evidentiary support, future research should examine content knowledge as a standalone factor to better understand its role in persuasive writing proficiency. These exclusions represent a limitation of the current study but also highlight promising directions for advancing both theoretical and practical applications of automated writing assessment.

Sixth, when preparing the data for factor analysis, detailed information on student exclusions is lacking, particularly regarding students who receive Individualized Education Plans and English Learners status. Furthermore, the distribution of exclusions by grade level is not provided. It is possible that certain groups, such as 6th grade students, especially those with

IEPs, EL status, or low socioeconomic status, may have been excluded at higher rates. However, due to the absence of this specific data, the potential impact of these exclusions on the results cannot be fully evaluated. This limitation hinders our ability to assess the representativeness of the sample and the generalizability of the findings across different student populations.

Seventh, the fit statistics for the factor analysis were relatively marginal, suggesting that the model may not have fully captured the underlying data structure as intended. This limitation highlights the need for further refinement of the model, e.g., the incorporation of additional relevant features and a more careful consideration of measurement inclusion and variable selection. Such improvements may enhance model fit and better align the results with established benchmarks in the literature.

5.3 CONCLUSION

This dissertation contributes to the literature by providing empirical evidence of the features that are informative for evaluating persuasive writing quality and the potential of using content-generation tools (e.g., GPT) to provide feedback and revise student essays in alignment with these features. While our findings demonstrate the promise of AI-driven tools for offering individualized writing instruction to students, it is important to acknowledge potential limitations (Zhou et al., 2023). Nonetheless, the results highlight the transformative potential of AI in K-12 writing instruction, offering new opportunities for students to develop their writing skills. This study aims to present a viable solution and a starting point for preparing both teachers and students to harness the capabilities of emerging technologies in writing instruction and assessment. By doing so, it seeks to enhance learning outcomes and foster the growth of student writing proficiency.

BIBLIOGRAPHY

Abrahams, R. D. (1972). The training of the man of words in talking sweet. *Language in Society*, *1*(1), 15–29. https://doi.org/10.1017/S0047404500006527

Aldisert, R. J. (2009). *Opinion Writing*. AuthorHouse.

Alexander, P. A., Graham, S., & Harris, K. R. (1998). A Perspective on Strategy Research: Progress and Prospects. *Educational Psychology Review*, *10*(2), 129–154. https://doi.org/10.1023/A:1022185502996

Alward, A. S., Chua, C. M., & Bidin, S. J. (2012). Hedges and boosters in the Yemeni EFL undergraduates' persuasive essay: An empirical study. *The Internet Journal of Language, Culture and Society*, *34*, 1–12.

Anders, P. L., & Guzzetti, B. J. (2020). *Literacy Instruction in the Content Areas* (2nd ed.). Routledge. https://doi.org/10.4324/9781003064282

Andreev, L., & Uccelli, P. (2024). The secret life of connectives: A taxonomy to study individual differences in mid-adolescents' use of connectives in writing to persuade. *Reading and Writing*, *37*(1), 173–204. https://doi.org/10.1007/s11145-023-10425-3

Andrews, R., Torgerson, C., Low, G., & McGuinn, N. (2009). Teaching argument writing to 7- to 14-year-olds: An international review of the evidence of successful practice. *Cambridge Journal of Education*, *39*(3), 291–310. https://doi.org/10.1080/03057640903103751

Andrews, S., Veldre, A., & Clarke, I. E. (2020). Measuring Lexical Quality: The Role of Spelling Ability. *Behavior Research Methods*, *52*(6), 2257–2282. https://doi.org/10.3758/s13428-020-01387-3

Applebee, A. N. (1986). Chapter VI: Problems in Process Approaches: Toward a Reconceptualization of Process Instruction. *Teachers College Record*, *87*(6), 95–113. https://doi.org/10.1177/016146818608700606

Applebee, A. N., & et al., A. (1994). *NAEP 1992 Writing Report Card*. U. https://eric.ed.gov/?id=ED370119

Applebee, A. N., & Langer, J. A. (2015). *Writing Instruction That Works: Proven Methods for Middle and High School Classrooms*. Teachers College Press.

Armstrong, S. L., Williams, J. L., & Stahl, N. A. (2018). Reading and Writing. In *Handbook of College Reading and Study Strategy Research* (3rd ed.). Routledge.

Aryadoust, V., & Liu, S. (2015). Predicting EFL writing ability from levels of mental representation measured by Coh-Metrix: A structural equation modeling study. *Assessing Writing*, *24*, 35–58. https://doi.org/10.1016/j.asw.2015.03.001

Attard, K. (2012). The role of narrative writing in improving professional practice. *Educational Action Research*, *20*(1), 161–175. https://doi.org/10.1080/09650792.2012.647754

Aull, L. (2015). *First-Year University Writing: A Corpus-Based Study with Implications for Pedagogy*. Springer.

Aull, L. L., & Lancaster, Z. (2014). Linguistic Markers of Stance in Early and Advanced Academic Writing: A Corpus-based Comparison. *Written Communication*, *31*(2), 151–183. https://doi.org/10.1177/0741088314527055

Aziz, F. I. A., & Ahmad, U. K. (2017). Persuasive Writing: How Students Argue. *Sains Humanika*, *9*(4–2), Article 4–2. https://doi.org/10.11113/sh.v9n4-2.1356

Aziz, F. I. B. A., & Said, S. B. M. (2020). Developing a persuasive writing model for secondary school. *Educational Research for Policy and Practice*, *19*(2), 143–158. https://doi.org/10.1007/s10671-019-09253-6

Baffour, P., Saxberg, T., & Crossley, S. (2023). Analyzing Bias in Large Language Model Solutions for Assisted Writing Feedback Tools: Lessons from the Feedback Prize Competition Series. In E. Kochmar, J. Burstein, A. Horbach, R. Laarmann-Quante, N. Madnani, A. Tack, V. Yaneva, Z. Yuan, & T. Zesch (Eds.), *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)* (pp. 242–246). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.bea-1.21

Bailey, A. L. (2007). *The Language Demands of School: Putting Academic English to the Test*. Yale University Press.

Ball, D. L. (2000). Bridging Practices: Intertwining Content and Pedagogy in Teaching and Learning to Teach. *Journal of Teacher Education*, *51*(3), 241–247. https://doi.org/10.1177/0022487100051003013

Barab, S., & Squire, K. (2004). Design-Based Research: Putting a Stake in the Ground. In *Design-based Research*. Psychology Press.

Bateman, T. (2024). Teacher perspectives of ChatGPT as a pedagogical tool in the K-12 setting: A case study. *Quality Assurance in Education*, *33*(2), 203–217. https://doi.org/10.1108/QAE-02-2024-0042

Bauer, L. B. (2014). Concept Mapping: Developing Metacognitive Awareness in a Postsecondary Reading and Writing Classroom. *Journal of College Literacy & Learning*, *40*, 35–44.

Bazerman, C., & Paradis, J. G. (1991). *Textual Dynamics of the Professions: Historical and Contemporary Studies of Writing in Professional Communities*. Univ of Wisconsin Press.

Beers, S. F., & Nagy, W. E. (2009). Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre? *Reading and Writing*, *22*(2), 185–200. https://doi.org/10.1007/s11145-007-9107-5

Beers, S. F., & Nagy, W. E. (2011). Writing development in four genres from grades three to seven: Syntactic complexity and genre differentiation. *Reading and Writing*, *24*(2), 183–202. https://doi.org/10.1007/s11145-010-9264-9

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Benz, A., Kühnlein, P., & Stede, M. (2012). *Constraints in Discourse 3*. 1–237.

Berman, R. A., & Nir-Sagiv, B. (2004). Linguistic indicators of inter-genre differentiation in later language development. *Journal of Child Language*, *31*(2), 339–380. https://doi.org/10.1017/S0305000904006038

Berman, R., & Verhoeven, L. (2002). Cross-linguistic perspectives on the development of text-production abilities: Speech and writing. *Written Language & Literacy*, *5*(1), 1–43. https://doi.org/10.1075/wll.5.1.02ber

Berninger, V. W., Cartwright, A. C., Yates, C. M., Swanson, H. L., & Abbott, R. D. (1994). Developmental skills related to writing and reading acquisition in the intermediate grades. *Reading and Writing*, *6*(2), 161–196. https://doi.org/10.1007/BF01026911

Berninger, V. W., Vaughan, K., Abbott, R. D., Begay, K., Coleman, K. B., Curtin, G., Hawkins, J. M., & Graham, S. (2002). Teaching spelling and composition alone and together: Implications for the simple view of writing. *Journal of Educational Psychology*, *94*(2), 291–304. https://doi.org/10.1037/0022-0663.94.2.291

Berninger, V. W., & Winn, W. D. (2006). Implications of Advancements in Brain Research and Technology for Writing Development, Writing Instruction, and Educational Evolution. In *Handbook of writing research* (pp. 96–114). The Guilford Press.

Beyreli, L., & Konuk, S. (2018). A Research on The Improvement of Persuasive Writing Skill of Sixth Grade Students in Secondary School. *TED EĞİTİM VE BİLİM*, *43*(193). https://doi.org/10.15390/EB.2018.7520

Biber, D. (1989). *A typology of English texts*. *27*(1), 3–44. https://doi.org/10.1515/ling.1989.27.1.3

Bickenbach, J. E., & Davies, J. M. (1996). *Good Reasons for Better Arguments: An Introduction to the Skills and Values of Critical Thinking*. Broadview Press.

Bishop, A. E., Sawyer, M., Alber-Morgan, S. R., & Boggs, M. (2015). Effects of a Graphic Organizer Training Package on the Persuasive Writing of Middle School Students with Autism. *Education and Training in Autism and Developmental Disabilities*, *50*(3), 290–302.

Blumner, J., & Childers, P. B. (2016). *WAC Partnerships Between Secondary and Postsecondary Institutions*. Parlor Press LLC.

Bondie, R. S., Dahnke, C., & Zusho, A. (2019). How Does Changing "One-Size-Fits-All" to Differentiated Instruction Affect Teaching? *Review of Research in Education*, *43*(1), 336–362. https://doi.org/10.3102/0091732X18821130

Boscolo, P., & Mason, L. (2001). Writing to Learn, Writing to Transfer. In P. Tynjälä, L. Mason, & K. Lonka (Eds.), *Writing as a Learning Tool: Integrating Theory and Practice* (pp. 83–104). Springer Netherlands. https://doi.org/10.1007/978-94-010-0740-5_6

Bouwer, R., Béguin, A., Sanders, T., & van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. *Language Testing*, *32*(1), 83–100. https://doi.org/10.1177/0265532214542994

Bouwer, R., Koster, M., & van den Bergh, H. (2018). Effects of a strategy-focused instructional program on the writing quality of upper elementary students in the Netherlands. *Journal of Educational Psychology*, *110*(1), 58–71. https://doi.org/10.1037/edu0000206

Bowers, A. J. (2011). What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school. *Educational Research and Evaluation*, *17*(3), 141–159. https://doi.org/10.1080/13803611.2011.597112

Boyle, J. R., & Hindman, A. H. (2015). Scaffolding the persuasive writing of middle school students. *Middle Grades Research Journal*, *10*(3), 43–60.

Breznitz, Z. (1997). Effects of accelerated reading rate on memory for text among dyslexic readers. *Journal of Educational Psychology*, *89*(2), 289–297. https://doi.org/10.1037/0022-0663.89.2.289

Broadhead, G. J. (1999). Addressing multiple goals for engineering writing: The role of course-specific websites. *Language and Learning across the Disciplines*, *3*(2), 19–43.

Brockman, E. (2020). Reframing Writing Prompts to Foster Nuanced Arguments: To What Extent? *English Journal*, *109*(6), 37–44. https://doi.org/10.58680/ej202030783

Brown, A., & Green, T. (2014). Issues and Trends in Instructional Technology: Maximizing Budgets and Minimizing Costs in Order to Provide Personalized Learning Opportunities. In M. Orey, S. A. Jones, & R. M. Branch (Eds.), *Educational Media and Technology Yearbook: Volume 38* (pp. 11–24). Springer International Publishing. https://doi.org/10.1007/978-3-319-06314-0_2

Bryant, D. P., Goodwin, M., Bryant, B. R., & Higgins, K. (2003). Vocabulary Instruction for Students with Learning Disabilities: A Review of the Research. *Learning Disability Quarterly*, *26*(2), 117–128. https://doi.org/10.2307/1593594

Bujang, M. A., & Baharum, N. (2016). Sample Size Guideline for Correlation Analysis. *World Journal of Social Science Research*, *3*(1), 37. https://doi.org/10.22158/wjssr.v3n1p37

Bürgler, S., Hoyle, R. H., & Hennecke, M. (2021). Flexibility in using self-regulatory strategies to manage self-control conflicts: The role of metacognitive knowledge, strategy repertoire,

and feedback monitoring. *European Journal of Personality*, *35*(6), 861–880. https://doi.org/10.1177/0890207021992907

Burstein, J., Kukich, K., Wolff, S., & Lu, C. (1998, April). *Computer analysis of essays*.

Butterfuss, R., Roscoe, R. D., Allen, L. K., McCarthy, K. S., & McNamara, D. S. (2022). Strategy Uptake in Writing Pal: Adaptive Feedback and Instruction. *Journal of Educational Computing Research*, *60*(3), 696–721. https://doi.org/10.1177/07356331211045304

Byrne, B. M. (2016). Adaptation of Assessment Scales in Cross-National Research: Issues, Guidelines, and Caveats. *International Perspectives in Psychology*, *5*(1), 51–65. https://doi.org/10.1037/ipp0000042

Cain, K. (2007). Syntactic awareness and reading ability: Is there any evidence for a special relationship? *Applied Psycholinguistics*, *28*(4), 679–694. https://doi.org/10.1017/S0142716407070361

Calfee, R. C., & and Chambliss, M. J. (1987). The Structural Design Features of Large Texts. *Educational Psychologist*, *22*(3–4), 357–378. https://doi.org/10.1080/00461520.1987.9653057

Camfield, E. K. (2016). Mediated-Efficacy: Hope for "Helpless" Writers. *Journal of Developmental Education*, *39*(3), 2–11.

Carvalhais, L., Limpo, T., & Pereira, L. Á. (2021). The Contribution of Word-, Sentence-, and Discourse-Level Abilities on Writing Performance: A 3-Year Longitudinal Study. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.668139

Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*. https://doi.org/10.1207/s15327906mbr0102_10

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language Deficits in Poor Comprehenders: A Case for the Simple View of Reading. *Journal of Speech, Language, and Hearing Research*, *49*(2), 278–293. https://doi.org/10.1044/1092-4388(2006/023)

Cer, E. (2019). The Instruction of Writing Strategies: The Effect of the Metacognitive Strategy on the Writing Skills of Pupils in Secondary Education. *Sage Open*, *9*(2), 2158244019842681. https://doi.org/10.1177/2158244019842681

Chambliss, M. J. (1995). Text Cues and Strategies Successful Readers Use to Construct the Gist of Lengthy Written Arguments. *Reading Research Quarterly*, *30*(4), 778–807. https://doi.org/10.2307/748198

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, *20*(1), 43. https://doi.org/10.1186/s41239-023-00411-8

Chen, S., Wei, B., & Zhang, H. (2023). Exploring high school students' disciplinary science identities and their differences. *International Journal of Science and Mathematics Education*, *21*(2), 377–394. https://doi.org/10.1007/s10763-022-10257-7

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Cho, M., Kim, Y.-S. G., & Olson, C. B. (2021). Does perspective taking matter for writing? Perspective taking in source-based analytical writing of secondary students. *Reading and Writing*, *34*(8), 2081–2101. https://doi.org/10.1007/s11145-021-10136-7

Christian, S., & Iryna, G. (2014). Identifying Argumentative Discourse Structures in Persuasive Essays. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. https://doi.org/10.3115/v1/d14-1006

Christie, F. (1998). Learning the Literacies of Primary and Secondary Schooling. In *Literacy and Schooling*. Routledge.

Chuang, P.-L., & Yan, X. (2023). Connecting source use and argumentation in L2 integrated argumentative writing performance. *Journal of Second Language Writing*, *60*, 101003. https://doi.org/10.1016/j.jslw.2023.101003

Chuang, Y.-T. (2014). Increasing Learning Motivation and Student Engagement through the Technology- Supported Learning Environment. *Creative Education*, *05*(23), Article 23. https://doi.org/10.4236/ce.2014.523221

Clark, D. B., & Sampson, V. D. (2007). Personally-Seeded Discussions to Scaffold Online Argumentation. *International Journal of Science Education*, *29*(3), 253–277. https://doi.org/10.1080/09500690600560944

Cochran, K., Cohn, C., Hastings, P., Tomuro, N., & Hughes, S. (2024). Using BERT to Identify Causal Structure in Students' Scientific Explanations. *International Journal of Artificial Intelligence in Education*, *34*(3), 1248–1286. https://doi.org/10.1007/s40593-023-00373-y

Cochran, K., Cohn, C., Hutchins, N., Biswas, G., & Hastings, P. (2022). Improving Automated Evaluation of Formative Assessments with Text Data Augmentation. In M. M. Rodrigo, N. Matsuda, A. I. Cristea, & V. Dimitrova (Eds.), *Artificial Intelligence in Education* (pp. 390–401). Springer International Publishing. https://doi.org/10.1007/978-3-031-11644-5_32

Coirier, P., Andriessen, J., & Chanquoy, L. (1999). From planning to translating: The specificity of argumentative writing. *Foundations of Argumentative Text Processing*, 1–28.

Coker Jr., D. L., Jennings, A. S., Farley-Ripple, E., & MacArthur, C. A. (2018). The type of writing instruction and practice matters: The direct and indirect effects of writing instruction and student practice on reading achievement. *Journal of Educational Psychology*, *110*(4), 502–517. https://doi.org/10.1037/edu0000232

Conley, D. T. (2014). *The common core state standards: Insight into their development.* Washington, DC: Council of Chief State School Officers.

Connor, U. (2004). Intercultural rhetoric research: Beyond texts. *Journal of English for Academic Purposes*, *3*(4), 291–304. https://doi.org/10.1016/j.jeap.2004.07.003

Connor, U., & Lauer, J. (1985). Understanding persuasive essay writing: Linguistic/rhetorical approach. *Text - Interdisciplinary Journal for the Study of Discourse*, *5*(4), 309–326. https://doi.org/10.1515/text.1.1985.5.4.309

Correnti, R., Matsumura, L. C., Wang, E., Litman, D., Rahimi, Z., & Kisa, Z. (2020). Automated Scoring of Students' Use of Text Evidence in Writing. *Reading Research Quarterly*, *55*(3), 493–520. https://doi.org/10.1002/rrq.281

Costello, A. B., & Osborne, J. (2019). *Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis*. https://doi.org/10.7275/JYJ1-4868

Crammond, J. G. (n.d.). *An analysis of argument structure in expert and student persuasive writing /*. McGill University. Retrieved February 15, 2025, from https://escholarship.mcgill.ca/concern/theses/wm117q58b

CRAMMOND, J. G. (1998). The Uses and Complexity of Argument Structures in Expert and Student Persuasive Writing. *Written Communication*, *15*(2), 230–268. https://doi.org/10.1177/0741088398015002004

Crossley, S. A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, *11*(3), 415–443. https://doi.org/10.17239/jowr-2020.11.03.01

Crossley, S. A., Baffour, P., Tian, Y., Picou, A., Benner, M., & Boser, U. (2022). The persuasive essays for rating, selecting, and understanding argumentative and discourse elements (PERSUADE) corpus 1.0. *Assessing Writing*, *54*, 100667. https://doi.org/10.1016/j.asw.2022.100667

Crossley, S. A., & Kyle, K. (2018). Analyzing Spoken and Written Discourse: A Role for Natural Language Processing Tools. In A. Phakiti, P. De Costa, L. Plonsky, & S. Starfield (Eds.), *The Palgrave Handbook of Applied Linguistics Research Methodology* (pp. 567–594). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-59900-1_25

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, *35*(2), 115–135. https://doi.org/10.1111/j.1467-9817.2010.01449.x

Crossley, S. A., & McNamara, D. S. (2016). Say more and be more coherent: How text elaboration and cohesion can increase writing quality. *Journal of Writing Research*, *7*(3), 351–370. https://doi.org/10.17239/jowr-2016.07.03.02

Crossley, S. A., Yang, H. S., & McNamara, D. S. (2014). What's so Simple about Simplified Texts? A Computational and Psycholinguistic Investigation of Text Comprehension and Text Processing. *Reading in a Foreign Language*, *26*(1), 92–113.

Crossley, S., Heintz, A., Choi, J. S., Batchelor, J., Karimi, M., & Malatinszky, A. (2023). A large-scaled corpus for assessing text readability. *Behavior Research Methods*, *55*(2), 491–507. https://doi.org/10.3758/s13428-022-01802-x

Crowhurst, M. (1980). Syntactic Complexity and Teachers' Quality Ratings of Narrations and Arguments. *Research in the Teaching of English*, *14*(3), 223–231. https://doi.org/10.58680/rte198015796

Crowhurst, M. (1990). Teaching and Learning the Writing of Persuasive/Argumentative Discourse. *Canadian Journal of Education / Revue Canadienne de l'éducation*, *15*(4), 348–359. https://doi.org/10.2307/1495109

Crowhurst, M., & Piche, G. L. (1979). Audience and Mode of Discourse Effects on Syntactic Complexity in Writing at Two Grade Levels. *Research in the Teaching of English*, *13*(2), 101–109. https://doi.org/10.58680/rte197917847

Cumberworth, T. J., & Hunt, J. A. (1998). *Improving Middle School Student Writing Skills and Attitudes toward Writing*. https://eric.ed.gov/?id=ED420865

Cutler, L., & Graham, S. (2008). Primary grade writing instruction: A national survey. *Journal of Educational Psychology*, *100*(4), 907–919. https://doi.org/10.1037/a0012656

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of Reading Comprehension: Relative Contributions of Word Recognition, Language Proficiency, and Other Cognitive Skills Can Depend on How Comprehension Is Measured. *Scientific Studies of Reading*, *10*(3), 277–299. https://doi.org/10.1207/s1532799xssr1003_5

De La Paz, S., Ferretti, R., Wissinger, D., Yee, L., & MacArthur, C. (2012). Adolescents' Disciplinary Use of Evidence, Argumentative Strategies, and Organizational Structure in Writing About Historical Controversies. *Written Communication*, *29*(4), 412–454. https://doi.org/10.1177/0741088312461591

De La Paz, S., & Graham, S. (1997). Effects of dictation and advanced planning instruction on the composing of students with writing and learning problems. *Journal of Educational Psychology*, *89*(2), 203–222. https://doi.org/10.1037/0022-0663.89.2.203

De La Paz, S., & Graham, S. (2002). Explicitly teaching strategies, skills, and knowledge: Writing instruction in middle school classrooms. *Journal of Educational Psychology*, *94*(4), 687–698. https://doi.org/10.1037/0022-0663.94.4.687

De La Paz, S., & Sherman, C. K. (2013). Revising Strategy Instruction in Inclusive Settings: Effects for English Learners and Novice Writers. *Learning Disabilities Research & Practice*, *28*(3), 129–141. https://doi.org/10.1111/ldrp.12011

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing*, *18*(1), 7–24. https://doi.org/10.1016/j.asw.2012.10.002

Deane, P., Odendahl, N., Quinlan, T., Fowles, M., Welsh, C., & Bivens-Tatum, J. (2008). Cognitive Models of Writing: Writing Proficiency as a Complex Integrated Skill. *ETS Research Report Series*, *2008*(2), i–36. https://doi.org/10.1002/j.2333-8504.2008.tb02141.x

Demetriou, A., Kazi, S., Makris, N., & Spanoudis, G. (2020). Cognitive ability, cognitive self-awareness, and school performance: From childhood to adolescence. *Intelligence*, *79*, 101432. https://doi.org/10.1016/j.intell.2020.101432

Deng, Z., Uccelli, P., & Snow, C. (2022). Diversity of Advanced Sentence Structures (DASS) in writing predicts argumentative writing quality and receptive academic language skills of fifth-to-eighth grade students. *Assessing Writing*, *53*, 100649. https://doi.org/10.1016/j.asw.2022.100649

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* (No. arXiv:1810.04805). arXiv. https://doi.org/10.48550/arXiv.1810.04805

Dhurandhar, A. (2009). Writing the Other: An Exercise in Empathy. *Journal for Learning through the Arts*, *5*(1). https://doi.org/10.21977/D95110039

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, *3*(3), 231–264. https://doi.org/10.1007/s11409-008-9029-x

Dobbs, C. L. (2014). Signaling organization and stance: Academic language use in middle grade persuasive writing. *Reading and Writing*, *27*(8), 1327–1352. https://doi.org/10.1007/s11145-013-9489-5

Dobbs, C. L., & Kearns, D. (2016). Using new vocabulary in writing: Exploring how word and learner characteristics relate to the likelihood that writers use newly taught vocabulary. *Reading and Writing*, *29*(9), 1817–1843. https://doi.org/10.1007/s11145-016-9654-8

Dockrell, J. E., Lindsay, G., Connelly, V., & Mackie, C. (2007). Constraints in the Production of Written Text in Children with Specific Language Impairments. *Exceptional Children*, *73*(2), 147–164. https://doi.org/10.1177/001440290707300202

Doewes, A., Kurdhi, N. A., & Saxena, A. (2023, July 5). Evaluating Quadratic Weighted Kappa as the Standard Performance Metric for Automated Essay Scoring. In *Proceedings of the 16th International Conference on Educational Data Mining*. International Educational Data Mining Society. https://doi.org/10.5281/zenodo.8115784

Dressen-Hammouda, D. (2008). From novice to disciplinary expert: Disciplinary identity and genre mastery. *English for Specific Purposes*, *27*(2), 233–252. https://doi.org/10.1016/j.esp.2007.07.006

Drew, S. V., Olinghouse, N. G., Faggella-Luby, M., & Welsh, M. E. (2017). Framework for disciplinary writing in science Grades 6–12: A national survey. *Journal of Educational Psychology*, *109*(7), 935–955. https://doi.org/10.1037/edu0000186

Duff, D., & Brydon, M. (2020). Estimates of individual differences in vocabulary size in English: How many words are needed to 'close the vocabulary gap'? *Journal of Research in Reading*, *43*(4), 454–481. https://doi.org/10.1111/1467-9817.12322

Durik, A. M., Britt, M. A., Reynolds, R., & Storey, J. (2008). The Effects of Hedges in Persuasive Arguments: A Nuanced Analysis of Language. *Journal of Language and Social Psychology*, *27*(3), 217–234. https://doi.org/10.1177/0261927X08317947

DURST, R., LAINE, C., SCHULTZ, L. M., & VILTER, W. (1990). Appealing Texts: The Persuasive Writing of High School Students. *Written Communication*, *7*(2), 232–255. https://doi.org/10.1177/0741088390007002003

Duschl, R. A., & Osborne, J. (2002). Supporting and Promoting Argumentation Discourse in Science Education. *Studies in Science Education*, *38*(1), 39–72. https://doi.org/10.1080/03057260208560187

Edwards, L. (2003). Writing Instruction in Kindergarten: Examining an Emerging Area of Research for Children with Writing and Reading Difficulties. *Journal of Learning Disabilities*, *36*(2), 136–148. https://doi.org/10.1177/00222194030360020

Eemeren, F. H. van, Henkemans, A. F. S., Eemeren, F. H. van, Grootendorst, R., & Henkemans, A. F. S. (2001). *Argumentation: Analysis, Evaluation, Presentation*. Routledge. https://doi.org/10.4324/9781410602442

Ekholm, E., Zumbrunn, S., & DeBusk-Lane, M. (2018). Clarifying an Elusive Construct: A Systematic Review of Writing Attitudes. *Educational Psychology Review*, *30*(3), 827–856. https://doi.org/10.1007/s10648-017-9423-5

Elleman, A. M., Oslund, E. L., Griffin, N. M., & Myers, K. E. (2019). A Review of Middle School Vocabulary Interventions: Five Research-Based Recommendations for Practice. *Language, Speech, and Hearing Services in Schools*, *50*(4), 477–492. https://doi.org/10.1044/2019_LSHSS-VOIA-18-0145

ElSayary, A. (2024a). An investigation of teachers' perceptions of using ChatGPT as a supporting tool for teaching and learning in the digital era. *Journal of Computer Assisted Learning*, *40*(3), 931–945. https://doi.org/10.1111/jcal.12926

ElSayary, A. (2024b). Enhancing Education in the Digital Era: Leveraging ChatGPT for Transformative Learning. In *International Perspectives on Educational Administration using Educational Inquiry*. Routledge.

Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, *4*(2), 139–155. https://doi.org/10.1016/1060-3743(95)90004-7

Englert, C. S., Okolo, C. M., & Mariage, T. V. (2009). Informational writing across the curriculum. In *Instruction and assessment for struggling writers: Evidence-based practices* (pp. 132–161). The Guilford Press.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's Argument Pattern for studying science discourse. *Science Education*, *88*(6), 915–933. https://doi.org/10.1002/sce.20012

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, *20*(1), 57. https://doi.org/10.1186/s41239-023-00425-2

Eser, R., & Ayaz, H. (2021). The effects of creative writing activities on narrative text writing skills and advanced reading awareness. *Research in Pedagogy*, *11*(2), 639–660. https://doi.org/10.5937/IstrPed2102639E

Evagorou, M., Papanastasiou, E., & Vrikki, M. (2023). What do we really know about students' written arguments? Evaluating written argumentation skills. *European Journal of Science and Mathematics Education*, *11*(4), 615–634. https://doi.org/10.30935/scimath/13284

Ezza, E.-S. Y., Ageeb, A. A., Sirry, R. O., & Mubarak, E. (2020). The Significance of Teaching Academic Writing as a Discipline-Specific Skill. In *Teaching Academic Writing as a Discipline-Specific Skill in Higher Education* (pp. 1–22). IGI Global Scientific Publishing. https://doi.org/10.4018/978-1-7998-2265-3.ch001

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Fan, J., Knoch, U., & Bond, T. (2019). Application of Rasch measurement theory in language assessment: Using measurement to enhance language assessment research and practice. *Papers in Language Testing and Assessment*, *8*(2), iii–ix.

Fang, Z., & Schleppegrell, M. J. (2010). Disciplinary Literacies Across Content Areas: Supporting Secondary Reading Through Functional Language Analysis. *Journal of Adolescent & Adult Literacy*, *53*(7), 587–597. https://doi.org/10.1598/JAAL.53.7.6

Fernsten, L. A., & Reda, M. (2011). Helping students meet the challenges of academic writing. *Teaching in Higher Education*, *16*(2), 171–182. https://doi.org/10.1080/13562517.2010.507306

Ferretti, R. P., & Graham, S. (2019). Argumentative writing: Theory, assessment, and instruction. *Reading and Writing*, *32*(6), 1345–1357. https://doi.org/10.1007/s11145-019-09950-x

Ferretti, R. P., Lewis, W. E., & Andrews-Weckerly, S. (2009). Do goals affect the structure of students' argumentative writing strategies? *Journal of Educational Psychology*, *101*(3), 577–589. https://doi.org/10.1037/a0014702

Fey, M. E., Catts, H. W., Proctor-Williams, K., Tomblin, J. B., & Zhang, X. (2004). Oral and Written Story Composition Skills of Children With Language Impairment. *Journal of Speech, Language, and Hearing Research*, *47*(6), 1301–1318. https://doi.org/10.1044/1092-4388(2004/098)

Fitzgerald, J., & Shanahan, T. (2000). Reading and Writing Relations and Their Development. *Educational Psychologist*, *35*(1), 39–50. https://doi.org/10.1207/S15326985EP3501_5

FitzPatrick, E. R., & McKeown, D. (2021). Writing from Multiple Source Texts: SRSD for Fifth Grade Learners in Inclusive Settings. *Learning Disabilities Research & Practice*, *36*(3), 188–200. https://doi.org/10.1111/ldrp.12257

Flower, L. (1994). *The Construction of Negotiated Meaning: A Social Cognitive Theory of Writing*. SIU Press.

Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication*, *32*(4), 365–387. https://doi.org/10.2307/356600

Fox, E. (2009). The Role of Reader Characteristics in Processing and Learning From Informational Text. *Review of Educational Research*, *79*(1), 197–261. https://doi.org/10.3102/0034654308324654

Fry, S. W., & Villagomez, A. (2012). Writing to Learn: Benefits and Limitations. *College Teaching*, *60*(4), 170–175. https://doi.org/10.1080/87567555.2012.697081

Fuchs, K. (2023). Exploring the opportunities and challenges of NLP models in higher education: Is Chat GPT a blessing or a curse? *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1166682

Furey, W. M., Marcotte, A. M., Wells, C. S., & Hintze, J. M. (2017). The Effects of Supplemental Sentence-Level Instruction for Fourth-Grade Students Identified as Struggling Writers. *Reading & Writing Quarterly*, *33*(6), 563–578. https://doi.org/10.1080/10573569.2017.1288591

Garwood, J. D., & Van Loan, C. L. (2019). Pre-service educators' dispositions toward inclusive practices for students with emotional and behavioural difficulties. *International Journal of Inclusive Education*, *23*(12), 1332–1347. https://doi.org/10.1080/13603116.2018.1447614

Gillam, R. B., & Johnston, J. R. (1992). Spoken and Written Language Relationships in Language/Learning-Impaired and Normally Achieving School-Age Children. *Journal of Speech, Language, and Hearing Research*, *35*(6), 1303–1315. https://doi.org/10.1044/jshr.3506.1303

Giri, V., & Paily, M. U. (2020). Effect of Scientific Argumentation on the Development of Critical Thinking. *Science & Education*, *29*(3), 673–690. https://doi.org/10.1007/s11191-020-00120-y

Gleason, M. M. . (1999). The Role of Evidence in Argumentative Writing. *Reading & Writing Quarterly*, *15*(1), 81–106. https://doi.org/10.1080/105735699278305

Glorfeld, L. W. (1995). An Improvement on Horn's Parallel Analysis Methodology for Selecting the Correct Number of Factors to Retain. *Educational and Psychological Measurement*, *55*(3), 377–393. https://doi.org/10.1177/0013164495055003002

Golley, A. M. (2015). *Teaching Writing for Students with Learning Disabilities in an Inclusive Classroom Setting: A Curriculum Development Project*. http://hdl.handle.net/1951/72988

Gómez Vera, G., Sotomayor, C., Bedwell, P., Domínguez, A. M., & Jéldrez, E. (2016). Analysis of lexical quality and its relation to writing quality for 4th grade, primary school students in Chile. *Reading and Writing*, *29*(7), 1317–1336. https://doi.org/10.1007/s11145-016-9637-9

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing Multilevel Analyses of Text Characteristics. *Educational Researcher*, *40*(5), 223–234. https://doi.org/10.3102/0013189X11413260

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 193–202. https://doi.org/10.3758/BF03195564

Graham, S. (2006). Strategy Instruction and the Teaching of Writing: A Meta-Analysis. In *Handbook of writing research* (pp. 187–207). The Guilford Press.

Graham, S. (2018). A Revised Writer(s)-Within-Community Model of Writing. *Educational Psychologist*, *53*(4), 258–279. https://doi.org/10.1080/00461520.2018.1481406

Graham, S. (2023). Writer(s)-within-Community Model of Writing as a Lens for Studying the Teaching of Writing. In *The Routledge International Handbook of Research on Writing* (2nd ed.). Routledge.

Graham, S., & Alves, R. A. (2021). Research and teaching writing. *Reading and Writing*, *34*(7), 1613–1621. https://doi.org/10.1007/s11145-021-10188-9

Graham, S., Capizzi, A., Harris, K. R., Hebert, M., & Morphy, P. (2014). Teaching writing to middle school students: A national survey. *Reading and Writing*, *27*(6), 1015–1042. https://doi.org/10.1007/s11145-013-9495-7

Graham, S., & Harris, K. (2010). Writing Difficulties. In *Handbook of Reading Disability Research*. Routledge.

Graham, S., & Harris, K. R. (1997). It Can Be Taught, But It Does Not Develop Naturally: Myths and Realities In Writing Instruction. *School Psychology Review*, *26*(3), 414–424. https://doi.org/10.1080/02796015.1997.12085875

Graham, S., & Harris, K. R. (2013). Common Core State Standards, Writing, and Students with LD: Recommendations. *Learning Disabilities Research & Practice*, *28*(1), 28–37. https://doi.org/10.1111/ldrp.12004

Graham, S., Harris, K. R., MacArthur, C. A., & Schwartz, S. (1991). Writing and Writing Instruction for Students with Learning Disabilities: Review of a Research Program. *Learning Disability Quarterly*, *14*(2), 89–114. https://doi.org/10.2307/1510517

Graham, S., Harris, K. R., & Mason, L. (2005). Improving the writing performance, knowledge, and self-efficacy of struggling young writers: The effects of self-regulated strategy development. *Contemporary Educational Psychology*, *30*(2), 207–241. https://doi.org/10.1016/j.cedpsych.2004.08.001

Graham, S., Hebert, M., Paige Sandbank, M., & Harris, K. R. (2016). Assessing the Writing Achievement of Young Struggling Writers: Application of Generalizability Theory. *Learning Disability Quarterly*, *39*(2), 72–82. https://doi.org/10.1177/0731948714555019

Graham, S., Kim, Y.-S., Cao, Y., Lee, J. won, Tate, T., Collins, P., Cho, M., Moon, Y., Chung, H. Q., & Olson, C. B. (2023). A meta-analysis of writing treatments for students in grades 6–12. *Journal of Educational Psychology*, *115*(7), 1004–1027. https://doi.org/10.1037/edu0000819

Graham, S., Kiuhara, S. A., & MacKay, M. (2020). The Effects of Writing on Learning in Science, Social Studies, and Mathematics: A Meta-Analysis. *Review of Educational Research*, *90*(2), 179–226. https://doi.org/10.3102/0034654320914744

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology*, *104*(4), 879–896. https://doi.org/10.1037/a0029185

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, *99*(3), 445–476. https://doi.org/10.1037/0022-0663.99.3.445

Graham, S., & Sandmel, K. (2011). The Process Writing Approach: A Meta-analysis. *The Journal of Educational Research*, *104*(6), 396–407. https://doi.org/10.1080/00220671.2010.488703

Graybeal, A., Sexton, J. D., & Pennebaker, J. W. (2002). The Role of Story-Making in Disclosure Writing: The Psychometrics of Narrative. *Psychology & Health*, *17*(5), 571–581. https://doi.org/10.1080/08870440290025786

Gregory, M. S.-J., & Lodge, J. M. (2015). Academic workload: The silent barrier to the implementation of technology-enhanced learning strategies in higher education. *Distance Education*, *36*(2), 210–230. https://doi.org/10.1080/01587919.2015.1055056

Grimberg, B. I., & Hand, B. (2009). Cognitive Pathways: Analysis of students' written texts for science understanding. *International Journal of Science Education*, *31*(4), 503–521. https://doi.org/10.1080/09500690701704805

Groves, F. H. (2016). A Longitudinal Study of Middle and Secondary Level Science Textbook Vocabulary Loads. *School Science and Mathematics*, *116*(6), 320–325. https://doi.org/10.1111/ssm.12183

Güneş, F. (2016). *Türkçe öğretimi: Yaklaşımlar ve modeller* (4. baskı). Pegem Akademi.

Halliday, M. A. K. (1995). Systemic Theory. In E. F. K. Koerner & R. E. Asher (Eds.), *Concise History of the Language Sciences* (pp. 272–276). Pergamon. https://doi.org/10.1016/B978-0-08-042580-1.50048-9

Hall-Mills, S., & Apel, K. (2015). Linguistic Feature Development Across Grades and Genre in Elementary Writing. *Language, Speech, and Hearing Services in Schools*, *46*(3), 242–255. https://doi.org/10.1044/2015_LSHSS-14-0043

Hall-Mills, S. S. (2009). *Linguistic feature development in elementary writing: Analysis of microstructure and macrostructure features in a narrative and an expository genre* [Ph.D., The Florida State University]. https://www.proquest.com/docview/849715588/abstract/EEAE26DD5254069PQ/1

Hamilton, C. L. (2011). *Anthem Guide to Essay Writing*. Anthem Press.

Hannah, L., Jang, E. E., Shah, M., & Gupta, V. (2023). Validity Arguments for Automated Essay Scoring of Young Students' Writing Traits. *Language Assessment Quarterly*, *20*(4–5), 399–420. https://doi.org/10.1080/15434303.2023.2288253

Harb, G. (2018). THE APPLICATION OF COCA CORPUS FOR A MORE SOPHISTICATED WORD CHOICE AND A BETTER EFL WRITING QUALITY. *INTED2018 Proceedings*, 5530–5537. 12th International Technology, Education and Development Conference. https://doi.org/10.21125/inted.2018.1306

Harmon, J. M., Hedrick, W. B., Wood, K. D., & Gress, M. (2005). Vocabulary Self-Selection: A Study of Middle-School Students' Word Selections from Expository Texts. *Reading Psychology*, *26*(3), 313–333. https://doi.org/10.1080/02702710590969689

Hartman, H. J. (Ed.). (2001). *Metacognition in Learning and Instruction* (Vol. 19). Springer Netherlands. https://doi.org/10.1007/978-94-017-2243-8

Hasan, R. (2014). Text in the Systemic-Functional Model. In W. U. Dressler (Ed.), *Current Trends in Textlinguistics* (pp. 228–246). De Gruyter. https://doi.org/10.1515/9783110853759.228

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hatzivassiloglou, V., & McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. *Proceedings of the 35th Annual Meeting on Association for Computational Linguistics -*, 174–181. https://doi.org/10.3115/976909.979640

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In *The science of writing: Theories, methods, individual differences, and applications* (pp. 1–27). Lawrence Erlbaum.

Hayes, J. R. (2012). Modeling and Remodeling Writing. *Written Communication*, *29*(3), 369–388. https://doi.org/10.1177/0741088312451260

Heider, K. (2009). *Information Literacy: The Missing Link in Early Childhood Education*. 3263–3269. https://www.learntechlib.org/primary/p/31150/

Hemberger, L., Kuhn, D., Matos, F., & Shi, Y. (2017). A Dialogic Path to Evidence-Based Argumentive Writing. *Journal of the Learning Sciences*, *26*(4), 575–607. https://doi.org/10.1080/10508406.2017.1336714

Hillocks, G. (2002). *The Testing Trap: How State Writing Assessments Control Learning*. Teachers College Press.

Hinkel, E. (2003). Simplicity Without Elegance: Features of Sentences in L1 and L2 Academic Texts. *TESOL Quarterly*, *37*(2), 275–301. https://doi.org/10.2307/3588505

Ho, V., & Li, C. (2018). The use of metadiscourse and persuasion: An analysis of first year university students' timed argumentative essays. *Journal of English for Academic Purposes*, *33*, 53–68. https://doi.org/10.1016/j.jeap.2018.02.001

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Howard, M. C. (2016). A Review of Exploratory Factor Analysis Decisions and Overview of Current Practices: What We Are Doing and How Can We Improve? *International Journal of Human–Computer Interaction*, *32*(1), 51–62. https://doi.org/10.1080/10447318.2015.1087664

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168–177. https://doi.org/10.1145/1014052.1014073

Hübner, S., Nückles, M., & Renkl, A. (2010). Writing learning journals: Instructional support to overcome learning-strategy deficits. *Learning and Instruction*, *20*(1), 18–29. https://doi.org/10.1016/j.learninstruc.2008.12.001

Hung, I. W., & Wyer, R. S. (2014). Effects of Self-Relevant Perspective-Taking on the Impact of Persuasive Appeals. *Personality and Social Psychology Bulletin*, *40*(3), 402–414. https://doi.org/10.1177/0146167213513474

Huot, B. (1990). The Literature of Direct Writing Assessment: Major Concerns and Prevailing Trends. *Review of Educational Research*, *60*(2), 237–263. https://doi.org/10.3102/00346543060002237

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216–225. https://doi.org/10.1609/icwsm.v8i1.14550

Ibrahim, R., Taha, H. Y., Dabous, A. A., & Khateb, A. (2013). Electronic Reading and Writing in Spoken and Written Arabic: A Case Study. *Theory and Practice in Language Studies*, *3*(9), 1497–1508. https://doi.org/10.4304/tpls.3.9.1497-1508

Ildikó Berzlánovich, & Gisela Redeker. (2012). Genre-dependent interaction of coherence and lexical cohesion in written discourse. *Corpus Linguistics and Linguistic Theory*, *8*(1), 183–208. https://doi.org/10.1515/cllt-2012-0008

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, *25*(3), 385–402. https://doi.org/10.1177/0265532208090158

Islam, J., Xiao, L., & Mercer, R. E. (2020). A Lexicon-Based Approach for Detecting Hedges in Informal Text. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 3109–3113). European Language Resources Association. https://aclanthology.org/2020.lrec-1.380/

Jagaiah, T. (2017). Analysis of Syntactic Complexity and Its Relationship to Writing Quality in Argumentative Essays. *Doctoral Dissertations*. https://digitalcommons.lib.uconn.edu/dissertations/1571

Jagaiah, T., Olinghouse, N. G., & Kearns, D. M. (2020). Syntactic complexity measures: Variation by genre, grade-level, students' writing abilities, and writing quality. *Reading and Writing*, *33*(10), 2577–2638. https://doi.org/10.1007/s11145-020-10057-x

Janda, H. K., Pawar, A., Du, S., & Mago, V. (2019). Syntactic, Semantic and Sentiment Analysis: The Joint Effect on Automated Essay Evaluation. *IEEE Access*, *7*, 108486–108503. IEEE Access. https://doi.org/10.1109/ACCESS.2019.2933354

Jeffery, J. V. (2009). Constructs of writing proficiency in US state and national writing assessments: Exploring variability. *Assessing Writing*, *14*(1), 3–24. https://doi.org/10.1016/j.asw.2008.12.002

Jeong, J., Gaffney, J. S., & Choi, J.-O. (2010). Availability and Use of Informational Texts in Second-, Third-, and Fourth-Grade Classrooms. *Research in the Teaching of English*, *44*(4), 435–456. https://doi.org/10.58680/rte201010850

Jescovitch, L. N., Scott, E. E., Cerchiara, J. A., Merrill, J., Urban-Lurain, M., Doherty, J. H., & Haudek, K. C. (2021). Comparison of Machine Learning Performance Using Analytic and Holistic Coding Approaches Across Constructed Response Assessments Aligned to a Science Learning Progression. *Journal of Science Education and Technology*, *30*(2), 150–167. https://doi.org/10.1007/s10956-020-09858-0

Jo, C. W. (2022). Mapping adolescent literacy across L1 backgrounds: Linguistic and discourse features as predictors of persuasive essay quality. *System*, *104*, 102698. https://doi.org/10.1016/j.system.2021.102698

Johnson, C. S., & Gelfand, S. (2013). Self-assessment and writing quality. *Academic Research International*, *4*(4), 571.

Johnson, L. L. (2016). Writing 2.0: How English Teachers Conceptualize Writing with Digital Technologies. *English Education*, *49*(1), 28–62. https://doi.org/10.58680/ee201628788

Johnson, M. D. (2014). Does Planning Really Help?: Effectiveness of Planning in L2 Writing. *Journal of Second Language Teaching & Research*, *3*(1), Article 1.

Jonassen, D. H., & Kim, B. (2010). Arguing to learn and learning to argue: Design justifications and guidelines. *Educational Technology Research and Development*, *58*(4), 439–457. https://doi.org/10.1007/s11423-009-9143-8

Junaidah Januin, W. H. O. (2021). Analysing ESL Persuasive Essay Writing Using Toulmin's Model of Argument. *Psychology and Education Journal*, *58*(1), 1810–1821. https://doi.org/10.17762/pae.v58i1.1034

Kaldaras, L., & Haudek, K. C. (2022). Validation of automated scoring for learning progression-aligned Next Generation Science Standards performance assessments. *Frontiers in Education*, *7*. https://doi.org/10.3389/feduc.2022.968289

Kame'enui, E. J., & Baumann, J. F. (2012). *Vocabulary Instruction: Research to Practice*. Guilford Press.

Kamhi, A. G., & Catts, H. W. (1999). Language and reading: Convergences and divergences. *Language and Reading Disabilities*, 1–24.

Kamil, M. L., Mosenthal, P. B., Pearson, P. D., & Barr, R. (2014). *Handbook of Reading Research*. Routledge.

Kamimura, N. (2011). A Study of EFL Argumentative Strategies: Factors that Affect the Choice of Organizational Pattern. *Asian Englishes*, *14*(2), 56–73. https://doi.org/10.1080/13488678.2011.10801307

Karasinski, C. M. (2023). Microstructure and Macrostructure Measures of Written Narrative, Expository, and Persuasive Language Samples. *Communication Disorders Quarterly*, *44*(3), 152–162. https://doi.org/10.1177/15257401221111334

Karbach, J. (1987). Using Toulmin's Model of Argumentation. *Journal of Teaching Writing*, *6*(1), 81–91.

Kelly-Riley, D. (2017). Common Core State Standards Initiative for Writing Program Administrators. *Writing Program Administration*, *40*(2), 193–221.

Kent, S. C., & Wanzek, J. (2016). The Relationship Between Component Skills and Writing Quality and Production Across Developmental Levels: A Meta-Analysis of the Last 25 Years. *Review of Educational Research*, *86*(2), 570–601. https://doi.org/10.3102/0034654315619491

Khairuddin, Z., Ismayatim, W. F. A., Ismail, O., Rahmat, N. H., & Zamri, N. A. (2021). *Exploring Critical Thinking in Writing*. 67–72. https://doi.org/10.2991/assehr.k.211227.012

Kieft, M., Rijlaarsdam, G., & van den Bergh, H. (2008). An aptitude–treatment interaction approach to writing-to-learn. *Learning and Instruction*, *18*(4), 379–390. https://doi.org/10.1016/j.learninstruc.2007.07.004

Kilgannon, A. (2022). *Reading and Writing*. https://opentextbc.ca/provincialenglishabe/chapter/reading-and-writing/

Kim, H., & Ro, E. (2023a). Additive extensive reading and syntactic development in second language writing: Analyses of syntactic complexity and sophistication in young EFL learners' book reports. *Journal of Second Language Writing*, *61*, 101040. https://doi.org/10.1016/j.jslw.2023.101040

Kim, H., & Ro, E. (2023b). Assessment of sentence sophistication in L2 spoken production: Expansion of verbs and argument structure constructions. *System*, *119*, 103175. https://doi.org/10.1016/j.system.2023.103175

Kim, M., & Crossley, S. A. (2018). Modeling second language writing quality: A structural equation investigation of lexical, syntactic, and cohesive features in source-based and independent writing. *Assessing Writing*, *37*, 39–56. https://doi.org/10.1016/j.asw.2018.03.002

Kim, M., Tian, Y., & Crossley, S. A. (2021). Exploring the relationships among cognitive and linguistic resources, writing processes, and written products in second language writing. *Journal of Second Language Writing*, *53*, 100824. https://doi.org/10.1016/j.jslw.2021.100824

Kim, Y.-S. G., & Graham, S. (2022). Expanding the Direct and Indirect Effects Model of Writing (DIEW): Reading–writing relations, and dynamic relations as a function of measurement/dimensions of written composition. *Journal of Educational Psychology*, *114*(2), 215–238. https://doi.org/10.1037/edu0000564

Kim, Y.-S., Park, C. H., & Park, Y. (2013). Is academic language use a separate dimension in beginning writing? Evidence from Korean. *Learning and Individual Differences*, *27*, 8–15. https://doi.org/10.1016/j.lindif.2013.06.002

Klassen, R. (2002). Writing in Early Adolescence: A Review of the Role of Self-Efficacy Beliefs. *Educational Psychology Review*, *14*(2), 173–203. https://doi.org/10.1023/A:1014626805572

Klein, P. D. (2006). The Challenges of Scientific Literacy: From the viewpoint of second-generation cognitive science. *International Journal of Science Education*, *28*(2–3), 143–178. https://doi.org/10.1080/09500690500336627

Klimova, B. F. (2013). Developing Thinking Skills in the Course of Academic Writing. *Procedia - Social and Behavioral Sciences*, *93*, 508–511. https://doi.org/10.1016/j.sbspro.2013.09.229

Kline, P. (2015). *A Handbook of Test Construction (Psychology Revivals): Introduction to Psychometric Design*. Routledge. https://doi.org/10.4324/9781315695990

Kober, N., McMurrer, J., & Silva, M. R. (2011). *State Test Score Trends through 2008-09, Part 4: Is Achievement Improving and Are Gaps Narrowing for Title I Students?* Center on Education Policy. https://eric.ed.gov/?id=ED522866

Koutsoftas, A. D., & Gray, S. (2012). Comparison of Narrative and Expository Writing in Students With and Without Language-Learning Disabilities. *Language, Speech, and Hearing Services in Schools*, *43*(4), 395–409. https://doi.org/10.1044/0161-1461(2012/11-0018)

Koutsoftas, A. D., & Petersen, V. (2017). Written cohesion in children with and without language learning disabilities. *International Journal of Language & Communication Disorders*, *52*(5), 612–625. https://doi.org/10.1111/1460-6984.12306

Kuo, C.-H. (1995). Cohesion and Coherence in Academic Writing: From Lexical Choice To Organization. *RELC Journal*, *26*(1), 47–62. https://doi.org/10.1177/003368829502600103

Kuzi, S., Cope, W., Ferguson, D., Geigle, C., & Zhai, C. (2019). Automatic Assessment of Complex Assignments using Topic Models. *Proceedings of the Sixth (2019) ACM Conference on Learning @ Scale*, 1–10. https://doi.org/10.1145/3330430.3333615

Laist, R. (2021). "Good Writing": Defining It and Teaching It. *Transformative Dialogues: Teaching and Learning Journal*, *14*(2), Article 2. https://journals.psu.edu/td/article/view/1399

LAUFER, B., & NATION, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, *16*(3), 307–322. https://doi.org/10.1093/applin/16.3.307

Lavelle, E., Smith, J., & O'Ryan, L. (2002). The writing approaches of secondary students. *British Journal of Educational Psychology*, *72*(3), 399–418. https://doi.org/10.1348/000709902320634564

Leijten, M., & Van Waes, L. (2013). Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes. *Written Communication*, *30*(3), 358–392. https://doi.org/10.1177/0741088313491692

Link, S., Mehrzad, M., & Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, *35*(4), 605–634. https://doi.org/10.1080/09588221.2020.1743323

Liu, D., & Wan, F. (2020). What Makes Proficient Writers' Essays More Persuasive? A Toulmin Perspective. *International Journal of TESOL Studies*. https://doi.org/10.46451/ijts.2020.06.01

Liu, F., & Stapleton, P. (2014). Counterargumentation and the cultivation of critical thinking in argumentative writing: Investigating washback from a high-stakes test. *System*, *45*, 117–128. https://doi.org/10.1016/j.system.2014.05.005

Liu, M., Liu, L., & Liu, L. (2018). Group awareness increases student engagement in online collaborative writing. *The Internet and Higher Education*, *38*, 1–8. https://doi.org/10.1016/j.iheduc.2018.04.001

Liu, Z., He, X., Liu, L., Liu, T., & Zhai, X. (2023). Context Matters: A Strategy to Pre-train Language Model for Science Education. In N. Wang, G. Rebolledo-Mendez, V. Dimitrova, N. Matsuda, & O. C. Santos (Eds.), *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky* (pp. 666–674). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36336-8_103

Lloyd-Jones, R. (1977). The Politics of Research into the Teaching of Composition. *College Composition & Communication*, *28*(3), 218–222. https://doi.org/10.58680/ccc197716365

Lu, X. (2017). Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing*, *34*(4), 493–511. https://doi.org/10.1177/0265532217710675

Ludwig, S., Mayer, C., Hansen, C., Eilers, K., & Brandt, S. (2021). Automated Essay Scoring Using Transformer Models. *Psych*, *3*(4), Article 4. https://doi.org/10.3390/psych3040056

Maamuujav, U. (2021). Examining lexical features and academic vocabulary use in adolescent L2 students' text-based analytical essays. *Assessing Writing*, *49*, 100540. https://doi.org/10.1016/j.asw.2021.100540

Maamuujav, U. (2022). *Unpacking Adolescent Writers' Texts: A Systematic Investigation of the Language Features in the Academic Writing of Linguistically Diverse Students* [Ph.D., University of California, Irvine]. https://www.proquest.com/docview/2691509390/abstract/88DC49028E8C409BPQ/1

Maamuujav, U., Olson, C. B., & Chung, H. (2021). Syntactic and lexical features of adolescent L2 students' academic writing. *Journal of Second Language Writing*, *53*, 100822. https://doi.org/10.1016/j.jslw.2021.100822

MacArthur, C. A., Graham, S., & Fitzgerald, J. (2008). *Handbook of Writing Research*. Guilford Press.

MacArthur, C. A., Jennings, A., & Philippakos, Z. A. (2019). Which linguistic features predict quality of argumentative writing for college basic writers, and how do those features change with instruction? *Reading and Writing*, *32*(6), 1553–1574. https://doi.org/10.1007/s11145-018-9853-6

Mackie, C., & Dockrell, J. E. (2004). The Nature of Written Language Deficits in Children With SLI. *Journal of Speech, Language, and Hearing Research*, *47*(6), 1469–1483. https://doi.org/10.1044/1092-4388(2004/109)

Magalhães, A. L. (2020). Teaching how to develop an argument using the Toulmin model. *International Journal of Multidisciplinary and Current Educational Research (IJMCER)*, *2*(3), 1–7.

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 55–60. https://doi.org/10.3115/v1/P14-5010

Marcotte, C., & Stokowski, P. A. (2021). Place meanings and national parks: A rhetorical analysis of social media texts. *Journal of Outdoor Recreation and Tourism*, *35*, 100383. https://doi.org/10.1016/j.jort.2021.100383

Masrul, M., & Yuliani, S. (2023). Interpreting collaborative reasoning and instructional strategies in argumentative writing. *Journal of English Language Teaching and Learning (JETLE)*, *5*(1), Article 1.

McCaffrey, D. F., Zhang, M., & Burstein, J. (2022). Across Performance Contexts: Using Automated Writing Evaluation to Explore Student Writing. In *Grantee Submission* (Vol. 6, pp. 167–199). https://eric.ed.gov/?id=ED618142

McCarthy, K. S., Roscoe, R. D., Likens, A. D., & McNamara, D. S. (2019). Checking It Twice: Does Adding Spelling and Grammar Checkers Improve Essay Quality in an Automated Writing Tutor? In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *Artificial Intelligence in Education* (pp. 270–282). Springer International Publishing. https://doi.org/10.1007/978-3-030-23204-7_23

McCormack, R. (2014). Reading Mandela. *Linguistics and the Human Sciences*, *9*(2), 177–200. https://doi.org/10.1558/lhs.v9i2.177

McElligott, N. (2014). "If Someone told me to Write it, I Wouldn't be able to and now I can" The Impact of Sixth Class Children Working Collaboratively with Their Peers to Enhance Their

Persuasive Communication and Writing Skills. In *Language, Literacy and Literature: Reimagining Teaching and Learning* (p. 115).

McKeown, D., FitzPatrick, E., Ennis, R., & Potter, A. (2020). Writing Is Revising: Improving Student Persuasive Writing through Individualized Asynchronous Audio Feedback. *Education and Treatment of Children*, *43*(1), 35–48. https://doi.org/10.1007/s43494-020-00004-4

McKnight, K., O'Malley, K., Ruzic, R., Horsley, M. K., Franey, J. J., & Bassett, K. (2016). Teaching in a Digital Age: How Educators Use Technology to Improve Student Learning. *Journal of Research on Technology in Education*, *48*(3), 194–211. https://doi.org/10.1080/15391523.2016.1175856

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic Features of Writing Quality. *Written Communication*, *27*(1), 57–86. https://doi.org/10.1177/0741088309351547

McNamara, D. S., & Graesser, A. C. (2012). Coh-Metrix: An Automated Tool for Theoretical and Applied Natural Language Processing. In *Applied Natural Language Processing: Identification, Investigation and Resolution* (pp. 188–205). IGI Global Scientific Publishing. https://doi.org/10.4018/978-1-60960-741-8.ch011

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge University Press.

McNeill, K. L., & Krajcik, J. (2009). Synergy Between Teacher Practices and Curricular Scaffolds to Support Students in Using Domain-Specific and Domain-General Knowledge in Writing Arguments to Explain Phenomena. *Journal of the Learning Sciences*, *18*(3), 416–460. https://doi.org/10.1080/10508400903013488

Means, M. L., & Voss, J. F. (1996). Who Reasons Well? Two Studies of Informal Reasoning Among Children of Different Grade, Ability, and Knowledge Levels. *Cognition and Instruction*, *14*(2), 139–178. https://doi.org/10.1207/s1532690xci1402_1

Menary, R. (2007). Writing as thinking. *Language Sciences*, *29*(5), 621–632. https://doi.org/10.1016/j.langsci.2007.01.005

Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A., & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence*, *6*, 100199. https://doi.org/10.1016/j.caeai.2023.100199

Midgette, E., Haria, P., & MacArthur, C. (2008). The effects of content and audience awareness goals for revision on the persuasive essays of fifth- and eighth-grade students. *Reading and Writing*, *21*(1), 131–151. https://doi.org/10.1007/s11145-007-9067-9

Miller, R. T., Mitchell, T. D., & Pessoa, S. (2016). Impact of source texts and prompts on students' genre uptake. *Journal of Second Language Writing*, *31*, 11–24. https://doi.org/10.1016/j.jslw.2016.01.001

Ming, L. S. (2005). Reduction of teacher workload in a formative assessment environment through use of online technology. *2005 6th International Conference on Information Technology Based Higher Education and Training*, F4A/18-F4A/21. https://doi.org/10.1109/ITHET.2005.1560302

Mo, Y., & Troia, G. A. (2017a). Predicting students' writing performance on the NAEP from student- and state-level variables. *Reading and Writing*, *30*(4), 739–770. https://doi.org/10.1007/s11145-016-9698-9

Mo, Y., & Troia, G. A. (2017b). Similarities and differences in constructs represented by U.S. States' middle school writing tests and the 2007 National Assessment of Educational Progress writing assessment. *Assessing Writing*, *33*, 48–67. https://doi.org/10.1016/j.asw.2017.06.001

Mohamad, F., Anuar, N., & Shidrah Mat Daud, N. (2022). Factors Influencing Academic Writing Behaviour Among Undergraduates. *International Journal of English Language Education*, *10*(2), 56. https://doi.org/10.5296/ijele.v10i2.20453

Monroe, B. W., & Troia, G. A. (2006). Teaching Writing Strategies to Middle School Students With Disabilities. *The Journal of Educational Research*, *100*(1), 21–33. https://doi.org/10.3200/JOER.100.1.21-33

Mostafa, A. M. (2017). An evaluation of sentiment analysis and classification algorithms for Arabic textual data. *International Journal of Computer Application*, *158*(3), 29–36.

Muijs, D. (2011). Leadership and organisational performance: From research to prescription? *International Journal of Educational Management*, *25*(1), 45–60. https://doi.org/10.1108/09513541111100116

Myklebust, H. R. (1980). What Is the Future for LD? *Journal of Learning Disabilities*, *13*(9), 4–7. https://doi.org/10.1177/002221948001300901

Mytkowicz, P., Goss, D., & Steinberg, B. (2014). Assessing Metacognition as a Learning Outcome in a Postsecondary Strategic Learning Course. *Journal of Postsecondary Education and Disability*, *27*(1), 51–62.

Nagy, W. E., & Scott, J. A. (2000). Vocabulary processes. In *Handbook of reading research, Vol. III* (pp. 269–284). Lawrence Erlbaum Associates Publishers.

Nagy, W., & Townsend, D. (2012). Words as Tools: Learning Academic Vocabulary as Language Acquisition. *Reading Research Quarterly*, *47*(1), 91–108. https://doi.org/10.1002/RRQ.011

National Center for Education Statistics. (2012). *The nation's report card: Writing 2011 (NCES 2012–470)*.

Negretti, R. (2012). Metacognition in Student Academic Writing: A Longitudinal Study of Metacognitive Awareness and Its Relation to Task Perception, Self-Regulation, and Evaluation of Performance. *Written Communication*, *29*(2), 142–179. https://doi.org/10.1177/0741088312438529

Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science*, *37*(4), 375–401. https://doi.org/10.1007/s11251-008-9053-x

Nelson, N. W., & Van Meter, A. M. (2007). Measuring Written Language Ability in Narrative Samples. *Reading & Writing Quarterly*, *23*(3), 287–309. https://doi.org/10.1080/10573560701277807

Nelzén, A. (2018). *Aristotle on social media? : Investigating non-profit organizations' usage of persuasive language in their posts on Twitter and Facebook*. https://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-70001

Newell, G. E., Beach, R., Smith, J., & VanDerHeide, J. (2011). Teaching and Learning Argumentative Reading and Writing: A Review of Research. *Reading Research Quarterly*, *46*(3), 273–304. https://doi.org/10.1598/RRQ.46.3.4

Nippold, M. A., Hesketh, L. J., Duthie, J. K., & Mansfield, T. C. (2005). Conversational Versus Expository Discourse. *Journal of Speech, Language, and Hearing Research*, *48*(5), 1048–1064. https://doi.org/10.1044/1092-4388(2005/073)

Nippold, M. A., & Ward-Lonergan, J. M. (2010). Argumentative writing in pre-adolescents: The role of verbal reasoning. *Child Language Teaching and Therapy*, *26*(3), 238–248. https://doi.org/10.1177/0265659009349979

Nkhobo, T., & Chaka, C. (2023). Student-Written versus ChatGPT-Generated Discursive Essays: A Comparative Coh-Metrix Analysis of Lexical Diversity, Syntactic Complexity, and Referential Cohesion. *International Journal of Education and Development Using Information and Communication Technology*, *19*(3), 69–84.

Nussbaum, E. M., & Kardash, C. M. (2005). The Effects of Goal Instructions and Text on the Generation of Counterarguments During Writing. *Journal of Educational Psychology*, *97*(2), 157–169. https://doi.org/10.1037/0022-0663.97.2.157

O'Connor, R. E., Beach, K. D., Sanchez, V. M., Kim, J. J., Knight-Teague, K., Orozco, G., & Jones, B. T. (2019). Teaching Academic Vocabulary to Sixth-Grade Students With Disabilities. *Learning Disability Quarterly*, *42*(4), 231–243. https://doi.org/10.1177/0731948718821091

Olinghouse, N. G. (2008). Student- and instruction-level predictors of narrative writing in third-grade students. *Reading and Writing*, *21*(1), 3–26. https://doi.org/10.1007/s11145-007-9062-1

Olinghouse, N. G., & Graham, S. (2009). The relationship between the discourse knowledge and the writing performance of elementary-grade students. *Journal of Educational Psychology*, *101*(1), 37–50. https://doi.org/10.1037/a0013462

Olinghouse, N. G., Graham, S., & Gillespie, A. (2015). The relationship of discourse and topic knowledge to fifth graders' writing performance. *Journal of Educational Psychology*, *107*(2), 391–406. https://doi.org/10.1037/a0037549

Olinghouse, N. G., & Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing*, *26*(1), 45–65. https://doi.org/10.1007/s11145-012-9392-5

Olson, D. R., & Torrance, N. (2009). *The Cambridge Handbook of Literacy*. Cambridge University Press.

Osborne, J. W. (n.d.). *What is Rotating in Exploratory Factor Analysis?* https://doi.org/10.7275/HB2G-M060

Özer, Ö., & Yükselir, C. (2022). Digital Writing Practices in the Information Age: Possibilities and Implications for English Language Teachers and Learners. In G. Yangın-Ekşi, S. Akayoglu, & L. Anyango (Eds.), *New Directions in Technology for Writing Instruction* (pp. 211–232). Springer International Publishing. https://doi.org/10.1007/978-3-031-13540-8_11

Pajares, F. (2003). Self-Efficacy Beliefs, Motivation, and Achievement in Writing: A Review of the Literature. *Reading & Writing Quarterly*, *19*(2), 139–158. https://doi.org/10.1080/10573560308222

Pajares, F., & Cheong, Y. F. (2003). Achievement goal orientations in writing:a developmental perspective. *International Journal of Educational Research*, *39*(4), 437–455. https://doi.org/10.1016/j.ijer.2004.06.008

Pajares, F., & Johnson, M. J. (1994). Confidence and Competence in Writing: The Role of Self-Efficacy, Outcome Expectancy, and Apprehension. *Research in the Teaching of English*, *28*(3), 313–331. https://doi.org/10.58680/rte199415378

Pajares, F., Miller, M. D., & Johnson, M. J. (1999). Gender differences in writing self-beliefs of elementary school students. *Journal of Educational Psychology*, *91*(1), 50–61. https://doi.org/10.1037/0022-0663.91.1.50

Pajares, F., & Valiante, G. (1999). Grade Level and Gender Differences in the Writing Self-Beliefs of Middle School Students. *Contemporary Educational Psychology*, *24*(4), 390–405. https://doi.org/10.1006/ceps.1998.0995

Pajares, F., & Valiante, G. (2001). Gender Differences in Writing Motivation and Achievement of Middle School Students: A Function of Gender Orientation? *Contemporary Educational Psychology*, *26*(3), 366–381. https://doi.org/10.1006/ceps.2000.1069

Palumbo, A., & Sanacore, J. (2009). Helping Struggling Middle School Literacy Learners Achieve Success. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *82*(6), 275–280. https://doi.org/10.3200/TCHS.82.6.275-280

Panadero, E., & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review*, *35*, 100416. https://doi.org/10.1016/j.edurev.2021.100416

Paris, S. G., & Newman, R. S. (1990). Development Aspects of Self-Regulated Learning. *Educational Psychologist*, *25*(1), 87–102. https://doi.org/10.1207/s15326985ep2501_7

Parson, A. (2013). *Writing for understanding: The effect of using informational writing on student science achievement* [Ph.D., Capella University]. https://www.proquest.com/docview/1458436461/abstract/A071E108824945D9PQ/1

Paul, S.-A. S., & Clarke, P. J. (2016). A systematic review of reading interventions for secondary school students. *International Journal of Educational Research*, *79*, 116–127. https://doi.org/10.1016/j.ijer.2016.05.011

Peloghitis, J. (n.d.). Difficulties and strategies in argumentative writing: A qualitative analysis. *Transformation in Language Education*, 399–406.

Perkins, K., Brutten, S. R., & Dare, D. E. (1991). A concurrent validity study of an ESL indirect measure of writing. *International Journal of Applied Linguistics*, *1*(2), 139–149. https://doi.org/10.1111/j.1473-4192.1991.tb00011.x

Philippakos, Z. A. T., & FitzPatrick, E. (2018). A Proposed Tiered Model of Assessment in Writing Instruction: Supporting All Student-Writers. *Insights into Learning Disabilities*, *15*(2), 149–173.

Plakans, L., & Gebril, A. (2013). Using multiple texts in an integrated writing assessment: Source text use as a predictor of score. *Journal of Second Language Writing*, *22*(3), 217–230. https://doi.org/10.1016/j.jslw.2013.02.003

Polio, C. G. (1997). Measures of Linguistic Accuracy in Second Language Writing Research. *Language Learning*, *47*(1), 101–143. https://doi.org/10.1111/0023-8333.31997003

Poulton, A., & Eliens, S. (2022). Explaining transformer-based models for automatic short answer grading. *Proceedings of the 5th International Conference on Digital Technology in Education*, 110–116. https://doi.org/10.1145/3488466.3488479

Puranik, C. S., Lombardino, L. J., & Altmann, L. J. (2007). Writing through retellings: An exploratory study of language-impaired and dyslexic populations. *Reading and Writing*, *20*(3), 251–272. https://doi.org/10.1007/s11145-006-9030-1

Puranik, C. S., Lombardino, L. J., & Altmann, L. J. P. (2008). Assessing the Microstructure of Written Language Using a Retelling Paradigm. *American Journal of Speech-Language Pathology*, *17*(2), 107–120. https://doi.org/10.1044/1058-0360(2008/012)

Puranik, C. S., & Lonigan, C. J. (2014). Emergent Writing in Preschoolers: Preliminary Evidence for a Theoretical Framework. *Reading Research Quarterly*, *49*(4), 453–467. https://doi.org/10.1002/rrq.79

Qin, J., & Karabacak, E. (2010). The analysis of Toulmin elements in Chinese EFL university argumentative writing. *System*, *38*(3), 444–456. https://doi.org/10.1016/j.system.2010.06.012

Qin, W., Jia, R., & Ren, W. (2024). Pragmatic Competence in an Email Writing Task: Influences of Situation, L1 Background, and L2 Proficiency. *Written Communication*, *41*(4), 726–755. https://doi.org/10.1177/07410883241263543

Qin, W., & Uccelli, P. (2016). Same language, different functions: A cross-genre analysis of Chinese EFL learners' writing performance. *Journal of Second Language Writing*, *33*, 3–17. https://doi.org/10.1016/j.jslw.2016.06.001

Ramadhanti, D., & Yanda, D. P. (2021). Students' Metacognitive Awareness and Its Impact on Writing Skill. *International Journal of Language Education*, *5*(3), 193–206.

Ramage, J. D., & Bean, J. C. (1992). *Writing arguments: A rhetoric with readings* (2nd ed). Macmillan Pub. Co.

Rambe, A. N. (2017). *THE EFFECT OF STORYTELLING METHOD ON STUDENTS WRITING NARRATIVE TEXT ABILITY AT THE ELEVENTH GRADE STUDENTS OF MAN PANYABUNGAN*. 77–79. http://aisteel2017.unimed.ac.id/wp-content/uploads/2018/03/77-79.pdf

Ramirez, G., Shaw, S. T., & Maloney, E. A. (2018). Math Anxiety: Past Research, Promising Interventions, and a New Interpretation Framework. *Educational Psychologist*, *53*(3), 145–164. https://doi.org/10.1080/00461520.2018.1447384

Ramli, N. F., Talib, O., Hassan, S. A., & Manaf, U. K. A. (2020). Development and Validation of an Instrument to Measure STEM Teachers' Instructional Preparedness. *Asian Journal of University Education*, *16*(3), 193–206.

Ramos, K. (2019). A Genre-Based Approach to Teaching Argument Writing. In L. C. de Oliveira, K. M. Obenchain, R. H. Kenney, & A. W. Oliveira (Eds.), *Teaching the Content Areas to English Language Learners in Secondary Schools: English Language Arts, Mathematics, Science, and Social Studies* (pp. 49–63). Springer International Publishing. https://doi.org/10.1007/978-3-030-02245-7_4

Raphael, T., Englert, C. S., & Kirschner, B. W. (1988). *Acquisition of Expository Writing Skills*. University of Illinois at Urbana-Champaign.

Reed, D. K., Binning, K., Jemison, E. A., & DeSalle, N. (2023). High–Quality Formative Writing Assessment for Middle School Students in Tier 2 Literacy Interventions. *Learning Disabilities Research & Practice*, *38*(1), 70–79. https://doi.org/10.1111/ldrp.12300

Reeves, T., & McKenney, S. (2012). 7 Things You Should Know About Educational Design Research. *Educause 7 Things Series*.

Revelle, W. (2015). *"Package 'Psych.'* The Comprehensive R Archive Network; https://cran.r-project.org/web/packages/psych/psych.pdf.

Richards, S. A. (2013). *Measurement and features of persuasive writing in undergraduate students with and without written language disorders* [Ph.D., Bowling Green State University]. https://www.proquest.com/docview/1645957370/abstract/DDDFE96EEF8449F1PQ/1

Rideout, J. C. (2016). Ethos, Character, and Discoursal Self in Persuasive Legal Writing. *Legal Writing: The Journal of the Legal Writing Institute*, *21*, 19.

Rodgers, D. B., Reed, D. K., Houchins, D. E., & Aloe, A. M. (2020). The writing abilities of juvenile justice youths: A confirmatory factor analysis. *The Journal of Educational Research*, *113*(6), 438–451. https://doi.org/10.1080/00220671.2020.1854160

Rodríguez-Málaga, L., Cueli, M., & Rodríguez, C. (2021). Exploring the effects of strategy-focused instruction in writing skills of 4TH grade students. *Metacognition and Learning*, *16*(1), 179–205. https://doi.org/10.1007/s11409-020-09247-3

Rohmatillah, R. Y. (2014). *Pengaaruh jumlah Deposito Mudharabah terhadap jumlah pembiayaan Murabahah pada bank pembiayaan rakyat syari'ah Al-Wadi'ah Tasikmala* [Other, UIN Sunan Gunung Djati Bandung]. https://digilib.uinsgd.ac.id/73483/

Roscoe, R. D., Snow, E. L., Allen, L. K., & McNamara, D. S. (2015). Automated Detection of Essay Revising Patterns: Applications for Intelligent Feedback in a Writing Tutor. In *Grantee Submission* (Vol. 10, Issue 1, pp. 59–79). https://eric.ed.gov/?id=ED565460

Roscoe, R. D., Wilson, J., Johnson, A. C., & Mayra, C. R. (2017). Presentation, expectations, and experience: Sources of student perceptions of automated writing evaluation. *Computers in Human Behavior*, *70*, 207–221. https://doi.org/10.1016/j.chb.2016.12.076

Rose, D., & Martin, J. R. (2012). *Learning to write, reading to learn: Genre, knowledge and pedagogy in the Sydney school*. Equinox. https://cir.nii.ac.jp/crid/1130282271267069312

Rosenberg, S. (1987). *Advances in Applied Psycholinguistics: Volume 2, Reading, Writing, and Language Learning*. CUP Archive.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*, 1–36. https://doi.org/10.18637/jss.v048.i02

Rowe, D. W., & Wilson, S. J. (2015). The Development of a Descriptive Measure of Early Childhood Writing: Results From the Write Start! Writing Assessment. *Journal of Literacy Research*, *47*(2), 245–292. https://doi.org/10.1177/1086296X15619723

Royse, E. A., Manzanares, A. D., Wang, H., Haudek, K. C., Azzarello, C. B., Horne, L. R., Druckenbrod, D. L., Shiroda, M., Adams, S. R., Fairchild, E., Vincent, S., Anderson, S. W., & Romulo, C. (2024). FEW questions, many answers: Using machine learning to assess how students connect food–energy–water (FEW) concepts. *Humanities and Social Sciences Communications*, *11*(1), 1–18. https://doi.org/10.1057/s41599-024-03499-z

Rubiaee, A. A. M., Darus, S., & Bakar, N. A. (2019). The Effect of Writing Knowledge on EFL Students' Ability in Composing Argumentative Essays. *Arab World English Journal*, *10*(4), 263–287.

Runge, I., Lazarides, R., Rubach, C., Richter, D., & Scheiter, K. (2023). Teacher-reported instructional quality in the context of technology-enhanced teaching: The role of teachers' digital competence-related beliefs in empowering learners. *Computers & Education*, *198*, 104761. https://doi.org/10.1016/j.compedu.2023.104761

Russell, T. (1997). Teaching Teachers: How I Teach IS the Message. In *Teaching about Teaching*. Routledge.

Saddler, B., & Graham, S. (2007). The Relationship between Writing Knowledge and Writing Performance among More and Less Skilled Writers. *Reading & Writing Quarterly*, *23*(3), 231–247. https://doi.org/10.1080/10573560701277575

Saddler, B., & Preschern, J. (2007). Improving Sentence Writing Ability through Sentence-Combining Practice. *TEACHING Exceptional Children*, *39*(3), 6–11. https://doi.org/10.1177/004005990703900301

Sadeghi, K., & Dilmaghani, S. K. (2013). The Relationship between Lexical Diversity and Genre in Iranian EFL Learners' Writings. *Journal of Language Teaching and Research*, *4*(2), 328–334.

Salahu-Din, D., Persky, H., & Miller, J. (2008). *The nation's report card: Writing 2007 (NCES 2008–468)*. National Center for Education Statistics, Institute of Education Sciences, US Department of Education, Washington, DC.

Salas, N., & Silvente, S. (2020). The role of executive functions and transcription skills in writing: A cross-sectional study across 7 years of schooling. *Reading and Writing*, *33*(4), 877–905. https://doi.org/10.1007/s11145-019-09979-y

Sampson, V., Enderle, P., Grooms, J., & Witte, S. (2013). Writing to Learn by Learning to Write During the School Science Laboratory: Helping Middle and High School Students Develop Argumentative Writing Skills as They Learn Core Ideas. *Science Education*, *97*(5), 643–670. https://doi.org/10.1002/sce.21069

Sanders, T., & Schilperoord, J. (2006). Text structure as a window on the cognition of writing. In *Handbook of writing research*.

Sanders, T., & Wijk, C. van. (1996). Text analysis as a research tool: How hierarchical text structure contributes to the understanding of conceptual processes in writing. In *The Science of Writing*. Routledge.

Sandoval, W. A., & Millwood, K. A. (2005). The Quality of Students' Use of Evidence in Written Scientific Explanations. *Cognition and Instruction*, *23*(1), 23–55. https://doi.org/10.1207/s1532690xci2301_2

Sarmiento, C. M., Truckenmiller, A. J., Cho, E., & Wang, H. (2024). Academic language use in middle school informational writing. *British Journal of Educational Psychology*, *n/a*(n/a). https://doi.org/10.1111/bjep.12724

Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. In *Advances in applied psycholinguistics* (pp. 142–175).

Scardamalia, M., & Paris, P. (1985). The Function of Explicit Discourse Knowledge in the Development of Text Representations and Composing Strategies. *Cognition and Instruction*, *2*(1), 1–39. https://doi.org/10.1207/s1532690xci0201_1

Schmitz, B., & Perels, F. (2011). Self-monitoring of self-regulation during math homework behaviour using standardized diaries. *Metacognition and Learning*, *6*(3), 255–273. https://doi.org/10.1007/s11409-011-9076-6

Schunk, D. H., & Zimmerman, B. J. (2007). Influencing Children's Self-Efficacy and Self-Regulation of Reading and Writing Through Modeling. *Reading & Writing Quarterly*, *23*(1), 7–25. https://doi.org/10.1080/10573560600837578

Scott, C. M. (2009). A Case for the Sentence in Reading Comprehension. *Language, Speech, and Hearing Services in Schools*, *40*(2), 184–191. https://doi.org/10.1044/0161-1461(2008/08-0042)

Scott, C. M., & Windsor, J. (2000). General Language Performance Measures in Spoken and Written Narrative and Expository Discourse of School-Age Children With Language Learning Disabilities. *Journal of Speech, Language, and Hearing Research*, *43*(2), 324–339. https://doi.org/10.1044/jslhr.4302.324

SHANAHAN, T., & SHANAHAN, C. (2008). Teaching Disciplinary Literacy to Adolescents: Rethinking Content- Area Literacy. *Harvard Educational Review*, *78*(1), 40–59. https://doi.org/10.17763/haer.78.1.v62444321p602101

Shemwell, M. (2020). Implementing Writing in Content Areas: Teachers' Perceptions as Writing Instructors. *Theses and Dissertations*. https://csuepress.columbusstate.edu/theses_dissertations/394

Shen, M., & Troia, G. (2018a). Evidence-based practices to improve the expository writing performance of students with learning disabilities: Strategy instruction vs genre study. *Perspectives on Language and Literacy*, *44*(2), 10–16.

Shen, M., & Troia, G. A. (2018b). Teaching Children With Language-Learning Disabilities to Plan and Revise Compare–Contrast Texts. *Learning Disability Quarterly*, *41*(1), 44–61. https://doi.org/10.1177/0731948717701260

Shermis, M. D., Burstein, J., & Bursky, S. A. (2013). Introduction to Automated Essay Evaluation. In *Handbook of Automated Essay Evaluation*. Routledge.

Siekmann, L., Parr, J. M., & Busse, V. (2022). Structure and coherence as challenges in composition: A study of assessing less proficient EFL writers' text quality. *Assessing Writing*, *54*, 100672. https://doi.org/10.1016/j.asw.2022.100672

Simon, S. (2008). 'A Single Brushstroke', writing through translation: Anne Carson. In P. St-Pierre & P. C. Kar (Eds.), *In Translation – Reflections, Refractions, Transformations* (pp. 107–116). John Benjamins Publishing Company. https://doi.org/10.1075/btl.71.13sim

Singh, M. (2011). The place of social justice in higher education and social change discourses. *Compare: A Journal of Comparative and International Education*, *41*(4), 481–494. https://doi.org/10.1080/03057925.2011.581515

Slomp, D. H. (2012). Challenges in assessing the development of writing ability: Theories, constructs and methods. *Assessing Writing*, *17*(2), 81–91. https://doi.org/10.1016/j.asw.2012.02.001

Smith, M. R. (2014). *Advanced legal writing: Theories and strategies in persuasive writing*.

Smyth, J. M. (1998). Written emotional expression: Effect sizes, outcome types, and moderating variables. *Journal of Consulting and Clinical Psychology*, *66*(1), 174.

Snow, C. E., & Uccelli, P. (2009). The Challenge of Academic Language. In D. R. Olson & N. Torrance (Eds.), *The Cambridge Handbook of Literacy* (pp. 112–133). Cambridge University Press. https://doi.org/10.1017/CBO9780511609664.008

Soltyska, A. (2024). Student and Teacher Views on the Use of Online Language Support Tools in EFL Exam Preparation. In *Technology-Mediated Language Learning and Teaching* (pp. 243–274). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-2687-9.ch009

Spencer, T. D., & Petersen, D. B. (2018). Bridging Oral and Written Language: An Oral Narrative Language Intervention Study With Writing Outcomes. *Language, Speech, and Hearing Services in Schools*, *49*(3), 569–581. https://doi.org/10.1044/2018_LSHSS-17-0030

Spielhagen, F. R. (2015). *Middle Grades Research Journal: Volume 10 #3*. IAP.

Stapleton, C. D. (1997). *Basic Concepts and Procedures of Confirmatory Factor Analysis*. https://eric.ed.gov/?id=ED407416

Stapleton, C. M., Zhang, H., & Berman, J. S. (2021). The Event-Specific Benefits of Writing About a Difficult Life Experience. *Europe's Journal of Psychology*, *17*(1), 53–69. https://doi.org/10.5964/ejop.2089

Stapleton, P., & Wu, Y. (Amy). (2015). Assessing the quality of arguments in students' persuasive writing: A case study analyzing the relationship between surface structure and substance. *Journal of English for Academic Purposes*, *17*, 12–23. https://doi.org/10.1016/j.jeap.2014.11.006

Stein, N. L., & Bernas, R. (1999). The early emergence of argumentative knowledge and skill. *Foundations of Argumentative Text Processing*, *5*, 97–116.

Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, *91*, 101894. https://doi.org/10.1016/j.learninstruc.2024.101894

Streibel, K. A. (2014). *The secondary/postsecondary writing gap: Obeying standards, striving for authenticity*. Doctoral dissertation.

Subrahmanian, V. S., & Reforgiato, D. (2008). AVA: Adjective-Verb-Adverb Combinations for Sentiment Analysis. *IEEE Intelligent Systems*, *23*(4), 43–50. https://doi.org/10.1109/MIS.2008.57

Sundari, H., & Febriyanti, R. H. (2021). The Analysis of Indonesian EFL Argumentative Writing Using Toulmin's Model: The Structure and Struggles from the Learners. *Scope : Journal of English Language Teaching*, *5*(2), Article 2. https://doi.org/10.30998/scope.v5i2.8544

Sürücü, L., Yıkılmaz, İ., & Maşlakçı, A. (2024). Exploratory Factor Analysis (EFA) in Quantitative Researches and Practical Considerations. *Gümüşhane Üniversitesi Sağlık Bilimleri Dergisi*, *13*(2), Article 2. https://doi.org/10.37989/gumussagbil.1183271

Susilawati, S., Gailea, N., & Masrupi, M. (2019). The Effectiveness of Facebook Towards Student's Writing Recount Text and Vocabulary Mastery. *Bahtera*, *18*(2), 135–147. https://doi.org/10.21009/BAHTERA.182.03

Sussex, U. of. (2007). Issues of Causality in Children's Reading Comprehension: Jane Oakhill Kate Cain. In *Reading Comprehension Strategies*. Psychology Press.

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (Seventh edition). Pearson.

Taboada, A., & Guthrie, J. T. (2006). Contributions of Student Questioning and Prior Knowledge to Construction of Knowledge from Reading Information Text. *Journal of Literacy Research*, *38*(1), 1–35. https://doi.org/10.1207/s15548430jlr3801_1

Taheri, P., & Nazmi, R. (2021). Improving EFL Learners Argumentative Writing Ability: Teacher vs. Peer Scaffolding. *Teaching English Language*, *15*(2). https://doi.org/10.22132/tel.2021.143348

Tasya, M. A. (2022). *Students' Difficulties in Writing an Argumentative Essay* [bachelorThesis, Jakarta: FITK UIN Syarif Hidayatullah Jakarta]. https://repository.uinjkt.ac.id/dspace/handle/123456789/59169

Taylor, K. S., Lawrence, J. F., Connor, C. M., & Snow, C. E. (2019). Cognitive and linguistic features of adolescent argumentative writing: Do connectives signal more complex reasoning? *Reading and Writing*, *32*(4), 983–1007. https://doi.org/10.1007/s11145-018-9898-6

The Condition of Education. (2009). U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Thomas, J. A. (2014). *Meaning in Interaction: An Introduction to Pragmatics*. Routledge. https://doi.org/10.4324/9781315842011

Thompson, B. (2000). Ten commandments of structural equation modeling. In *Reading and understanding MORE multivariate statistics* (pp. 261–283). American Psychological Association.

Tindale, J. A. (2002). Variance in the Meaning of Time by Family Cycle, Period, Social Context, and Ethnicity. In W. E. Pentland, A. S. Harvey, M. P. Lawton, & M. A. McColl (Eds.), *Time Use Research in the Social Sciences* (pp. 155–168). Springer US. https://doi.org/10.1007/0-306-47155-8_7

Tomas, C., Whitt, E., Lavelle-Hill, R., & Severn, K. (2019). Modeling Holistic Marks With Analytic Rubrics. *Frontiers in Education*, *4*. https://doi.org/10.3389/feduc.2019.00089

Tömen, M. (2016). *The Relationship Between Vocabulary Size, Lexical Diversity, Lexical Density and Efl Writing Scores: A Cross-Sectional Study* [M.A., Anadolu University (Turkey)]. https://www.proquest.com/docview/2561900728/abstract/ECCCA9EC8DF242BCPQ/1

Torrance, M., Arrimada, M., & Gardner, S. (2021). Child-level factors affecting rate of learning to write in first grade. *British Journal of Educational Psychology*, *91*(2), e12390. https://doi.org/10.1111/bjep.12390

Tortorelli, L. S. (2020). Beyond first grade: Examining word, sentence, and discourse text factors associated with oral reading rate in informational text in second grade. *Reading and Writing*, *33*(1), 143–170. https://doi.org/10.1007/s11145-019-09956-5

Toulmin, S. E. (2003). *The Uses of Argument*. Cambridge University Press.

Toutanova, K., Mitchell, M., & Manning, C. D. (2003). Optimizing Local Probability Models for Statistical Parsing. In N. Lavrač, D. Gamberger, H. Blockeel, & L. Todorovski (Eds.),

Machine Learning: ECML 2003 (pp. 409–420). Springer. https://doi.org/10.1007/978-3-540-39857-8_37

Tran, L. T. (2007). Learners' Motivation and Identity in the Vietnamese EFL Writing Classroom. *English Teaching: Practice and Critique*, *6*(1), 151–163.

Trapman, M., van Gelderen, A., van Schooten, E., & Hulstijn, J. (2018). Writing proficiency level and writing development of low-achieving adolescents: The roles of linguistic knowledge, fluency, and metacognitive knowledge. *Reading and Writing*, *31*(4), 893–926. https://doi.org/10.1007/s11145-018-9818-9

Troia, G. (2014). *Evidence-based practices for writing instruction*.

Troia, G. A. (2002). Teaching Writing Strategies to Children With Disabilities: Setting Generalization as the Goal. *Exceptionality*, *10*(4), 249–269. https://doi.org/10.1207/S15327035EX1004_3

Troia, G. A. (2011). *Instruction and Assessment for Struggling Writers: Evidence-Based Practices*. Guilford Press.

Troia, G. A., Harbaugh, A. G., Shankland, R. K., Wolbers, K. A., & Lawrence, A. M. (2013). Relationships between writing motivation, writing activity, and writing performance: Effects of grade, sex, and ability. *Reading and Writing*, *26*(1), 17–44. https://doi.org/10.1007/s11145-012-9379-2

Troia, G. A., Lin, S. C., Cohen, S., & Monroe, B. W. (2011). A Year in the Writing Workshop. *The Elementary School Journal*, *112*(1), 155–182. https://doi.org/10.1086/660688

Troia, G. A., & Maddox, M. E. (2004). Writing Instruction in Middle Schools: Special and General Education Teachers Share Their Views and Voice Their Concerns. In *Students Who Are Exceptional and Writing Disabilities*. Routledge.

Troia, G. A., & Olinghouse, N. G. (2013). The Common Core State Standards and Evidence-Based Educational Practices: The Case of Writing. *School Psychology Review*, *42*(3), 343–357. https://doi.org/10.1080/02796015.2013.12087478

Troia, G. A., Shankland, R. K., & Wolbers, K. A. (2012). Motivation Research in Writing: Theoretical and Empirical Considerations. *Reading & Writing Quarterly*, *28*(1), 5–28. https://doi.org/10.1080/10573569.2012.632729

Troia, G. A., Shen, M., & Brandon, D. L. (2019). Multidimensional Levels of Language Writing Measures in Grades Four to Six. *Written Communication*, *36*(2), 231–266. https://doi.org/10.1177/0741088318819473

Troia, G., Wang, H., & Lawrence, F. R. (2022). Latent profiles of writing-related skills, knowledge, and motivation for elementary students and their relations to writing performance across multiple genres. *Contemporary Educational Psychology*, *71*, 102100. https://doi.org/10.1016/j.cedpsych.2022.102100

Truckenmiller, A. J., Park, J., Dabo, A., & Wu Newton, Y.-C. (2019). Academic Language Instruction for Students in Grades 4 Through 8: A Literature Synthesis. *Journal of Research on Educational Effectiveness*, *12*(1), 135–159. https://doi.org/10.1080/19345747.2018.1536773

Truckenmiller, A. J., & Petscher, Y. (2020). The role of academic language in written composition in elementary and middle school. *Reading and Writing*, *33*(1), 45–66. https://doi.org/10.1007/s11145-019-09938-7

Truckenmiller, A., Shen, M., & Sweet, L. E. (2021). The Role of Vocabulary and Syntax in Informational Written Composition in Middle School. *Reading and Writing*, *34*(4), 911–943. https://doi.org/10.1007/s11145-020-10099-1

Turcotte, N., & Mesa, M. (2024). An Exploration of Middle and High School Teachers' Perceptions of ChatGPT. In *Artificial Intelligence Applications in K-12*. Routledge.

Uccelli, P., Dobbs, C. L., & Scott, J. (2013). Mastering Academic Language: Organization and Stance in the Persuasive Writing of High School Students. *Written Communication*, *30*(1), 36–62. https://doi.org/10.1177/0741088312469013

Valero Haro, A., Noroozi, O., Biemans, H. J. A., Mulder, M., & Banihashem, S. K. (2024). How does the type of online peer feedback influence feedback quality, argumentative essay writing quality, and domain-specific learning? *Interactive Learning Environments*, *32*(9), 5459–5478. https://doi.org/10.1080/10494820.2023.2215822

Visser, T., Maaswinkel, T., Coenders, F., & McKenney, S. (2018). Writing Prompts Help Improve Expression of Conceptual Understanding in Chemistry. *Journal of Chemical Education*, *95*(8), 1331–1335. https://doi.org/10.1021/acs.jchemed.7b00798

Wachsmuth, H., Stede, M., El Baff, R., Al-Khatib, K., Skeppstedt, M., & Stein, B. (2018). Argumentation Synthesis following Rhetorical Strategies. In E. M. Bender, L. Derczynski, & P. Isabelle (Eds.), *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 3753–3765). Association for Computational Linguistics. https://aclanthology.org/C18-1318/

Waldvogel, D. A. (2013). The Relationships between Vocabulary Learning Strategies and Vocabulary Size among Adult Spanish Foreign Language Learners. *Journal of Language Teaching and Research*, *4*(2), 209–219. https://doi.org/10.4304/jltr.4.2.209-219

Walton, D. (2013). *Methods of Argumentation*. Cambridge University Press.

Wang, H., Haudek, K. C., Manzanares, A. D., Romulo, C. L., & Royse, E. A. (2024). *Extending a Pretrained Language Model (BERT) using an Ontological Perspective to Classify Students' Scientific Expertise Level from Written Responses*. Research Square. https://doi.org/10.21203/rs.3.rs-3879583/v1

Wang, H., Li, T., Haudek, K., Royse, E. A., Manzanares, M., Adams, S., Horne, L., & Romulo, C. (2023). Is ChatGPT a Threat to Formative Assessment in College-Level Science? An

Analysis of Linguistic and Content-Level Features to Classify Response Types. In T. Schlippe, E. C. K. Cheng, & T. Wang (Eds.), *Artificial Intelligence in Education Technologies: New Development and Innovative Practices* (pp. 171–185). Springer Nature. https://doi.org/10.1007/978-981-99-7947-9_13

Wang, H., & Troia, G. A. (2023a). How students' writing motivation, teachers' personal and professional attributes, and writing instruction impact student writing achievement: A two-level hierarchical linear modeling study. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1213929

Wang, H., & Troia, G. A. (2023b). Writing Quality Predictive Modeling: Integrating Register-Related Factors. *Written Communication*, *40*(4), 1070–1112. https://doi.org/10.1177/07410883231185287

Warschauer, M., & Grimes, D. (2008). Automated Writing Assessment in the Classroom. *Pedagogies: An International Journal*, *3*(1), 22–36. https://doi.org/10.1080/15544800701771580

Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). *The Affordances and Contradictions of AI-Generated Text for Writers of English as a Second or Foreign Language* (SSRN Scholarly Paper No. 4404380). Social Science Research Network. https://doi.org/10.2139/ssrn.4404380

Waters, H. S., & Schneider, W. (2009). *Metacognition, Strategy Use, and Instruction*. Guilford Press.

Weinstein, C. E., & Hume, L. M. (1998). *Study strategies for lifelong learning* (pp. vii, 111). American Psychological Association. https://doi.org/10.1037/10296-000

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). American Psychological Association. https://doi.org/10.1037/10222-009

Widdowson, H. G. (1983). Talking Shop: On Literature and ELT. *ELT Journal*, *37*(1), 30–35.

Wijekumar, K., Graham, S., Harris, K. R., Lei, P.-W., Barkel, A., Aitken, A., Ray, A., & Houston, J. (2019). The roles of writing knowledge, motivation, strategic behaviors, and skills in predicting elementary students' persuasive writing from source material. *Reading and Writing*, *32*(6), 1431–1457. https://doi.org/10.1007/s11145-018-9836-7

Wiliana, A., & Djajanegara, S. (2019). THE EFFECT OF VOCABULARY MASTERY AND GRAMMAR TOWARDS STUDENTS' WRITING SKILL IN NARRATIVE TEXT. *INFERENCE: Journal of English Language Teaching*, *2*(3), Article 3. https://doi.org/10.30998/inference.v2i3.5783

Williams, C., & Beam, S. (2019). Technology and writing: Review of research. *Computers & Education*, *128*, 227–242. https://doi.org/10.1016/j.compedu.2018.09.024

Williamson, G. L. (2006). Student Readiness for Postsecondary Endeavors. In *Online Submission*. https://eric.ed.gov/?id=ED491516

Wilson, J., Ahrendt, C., Fudge, E. A., Raiche, A., Beard, G., & MacArthur, C. (2021). Elementary teachers' perceptions of automated feedback and automated scoring: Transforming the teaching and learning of writing using automated writing evaluation. *Computers & Education*, *168*, 104208. https://doi.org/10.1016/j.compedu.2021.104208

Wilson, J., Roscoe, R., & Ahmed, Y. (2017). Automated formative writing assessment using a levels of language framework. *Assessing Writing*, *34*, 16–36. https://doi.org/10.1016/j.asw.2017.08.002

Wingate, U. (2012). 'Argument!' helping students understand what essay writing is about. *Journal of English for Academic Purposes*, *11*(2), 145–154. https://doi.org/10.1016/j.jeap.2011.11.001

Wissinger, D. R., & De La Paz, S. (2020). Effects of Discipline-Specific Strategy Instruction on Historical Writing Growth of Students With Writing Difficulties. *Journal of Learning Disabilities*, *53*(3), 199–212. https://doi.org/10.1177/0022219420904343

Witte, S. P., & Faigley, L. (1981). Coherence, Cohesion, and Writing Quality. *College Composition & Communication*, *32*(2), 189–204. https://doi.org/10.58680/ccc198115912

Wolf, K. A. (1997). *The effects of an international pen pal project on the written-language skills of students with language-learning disabilities* [M.S., Utah State University]. https://www.proquest.com/docview/304380761/abstract/97721C23CC6645AEPQ/1

Wolfe, C. R., Britt, M. A., & Butler, J. A. (2009). Argumentation Schema and the Myside Bias in Written Argumentation. *Written Communication*, *26*(2), 183–209. https://doi.org/10.1177/0741088309333019

Wright, E., Smith, R., Vernon, M., Wall, R., & White, L. (2021). Inclusive, Multi-Partner Co-Creation for the Teaching of Special Educational Needs and Disabilities in Higher Education. *Journal of University Teaching and Learning Practice*, *18*(7), 25–38.

Wu, Y., & Schunn, C. D. (2021). The Effects of Providing and Receiving Peer Feedback on Writing Performance and Learning of Secondary School Students. *American Educational Research Journal*, *58*(3), 492–526. https://doi.org/10.3102/0002831220945266

Wulff, P., Mientus, L., Nowak, A., & Borowski, A. (2023). Utilizing a Pretrained Language Model (BERT) to Classify Preservice Physics Teachers' Written Reflections. *International Journal of Artificial Intelligence in Education*, *33*(3), 439–466. https://doi.org/10.1007/s40593-022-00290-6

Yamson (Maed), E. P., & L. Borong (EdD), N. (2022). Metacognitive Strategies for Developing Writing Skills. *International Journal of Social Sciences and Humanities Invention*, *9*(11), 7395–7403. https://doi.org/10.18535/ijsshi/v9i011.08

Zhang, M. (2018). Collaborative writing in the EFL classroom: The effects of L1 and L2 use. *System*, *76*, 1–12. https://doi.org/10.1016/j.system.2018.04.009

Zhou, J., Lu, T., Mishra, S., Brahma, S., Basu, S., Luan, Y., Zhou, D., & Hou, L. (2023). *Instruction-Following Evaluation for Large Language Models* (No. arXiv:2311.07911). arXiv. https://doi.org/10.48550/arXiv.2311.07911

Zhu, X., Wu, H., & Zhang, L. (2022). Automatic Short-Answer Grading via BERT-Based Deep Neural Networks. *IEEE Transactions on Learning Technologies*, *15*(3), 364–375. IEEE Transactions on Learning Technologies. https://doi.org/10.1109/TLT.2022.3175537

Zimmerbaum, K. Z. (2014). *Will the Introduction of a Critical Questioning Technique and the Toulmin Model Improve the Argumentative Essay Writing Scores of Students in an Eighth Grade English Language Arts Class?* [M.A.C.I., Caldwell College]. https://www.proquest.com/docview/1535293246/abstract/F1C14AF0E290454EPQ/1

Zupanc, K., & Bosnić, Z. (2017). Automated essay evaluation with semantic analysis. *Knowledge-Based Systems*, *120*, 118–132. https://doi.org/10.1016/j.knosys.2017.01.006